

UC Berkeley

UC Berkeley Previously Published Works

Title

The threshold theorem for the $(4 + 1)$ -dimensional yang-mills equation: An overview of the proof

Permalink

<https://escholarship.org/uc/item/5fj3j5pt>

Journal

Bulletin of the American Mathematical Society, 56(2)

ISSN

0273-0979

Authors

Oh, SJ
Tataru, D

Publication Date

2018

DOI

10.1090/BULL/1640

Peer reviewed

THE THRESHOLD THEOREM FOR THE $(4 + 1)$ -DIMENSIONAL YANG–MILLS EQUATION: AN OVERVIEW OF THE PROOF

SUNG-JIN OH AND DANIEL TATARU

ABSTRACT. This article is devoted to the energy critical hyperbolic Yang–Mills system in the $(4 + 1)$ dimensional Minkowski space, which is considered by the authors in a sequence of four papers [23], [24], [25] and [26]. The final outcome of these papers is twofold: (i) the Threshold Theorem, which asserts that global well-posedness and scattering hold for all topologically trivial initial data with energy below twice the ground state energy, and (ii) the Dichotomy Theorem, which for larger data in arbitrary topological classes provides a choice of two outcomes, either a global, scattering solution or a soliton bubbling off. In the last case, the bubbling off phenomena can happen either (a) in finite time, triggering a finite time blow-up, or (b) in infinite time. Our goal here is to describe these results, and to provide an overview of the flow of ideas within their proofs in [23], [24], [25] and [26].

CONTENTS

1. Introduction	1
2. The caloric gauge	9
3. Energy dispersed caloric Yang–Mills waves	18
4. Large data, causality and the temporal gauge	24
5. To bubble or not to bubble	27
References	34

1. INTRODUCTION

1.1. Lie groups and Lie algebras. Let \mathbf{G} be a compact noncommutative Lie group and \mathfrak{g} its associated Lie algebra. We denote by $Ad(O)X = OXO^{-1}$ the action of \mathbf{G} on \mathfrak{g} by conjugation (i.e., the adjoint action), and by $ad(X)Y = [X, Y]$ the associated action of \mathfrak{g} , which is given by the Lie bracket. We introduce the notation $\langle X, Y \rangle$ for a bi-invariant inner product on \mathfrak{g} ,

$$\langle [X, Y], Z \rangle = \langle X, [Y, Z] \rangle, \quad X, Y, Z \in \mathfrak{g},$$

or equivalently

$$\langle X, Y \rangle = \langle Ad(O)X, Ad(O)Y \rangle, \quad X, Y \in \mathfrak{g}, \quad O \in \mathbf{G}.$$

If \mathbf{G} is semisimple then one can take $\langle X, Y \rangle = -\text{tr}(ad(X)ad(Y))$ i.e. negative of the Killing form on \mathfrak{g} , which is then positive definite. However, a bi-invariant inner product on \mathfrak{g} exists for any compact Lie group G .

1.2. The Yang–Mills evolution. Let \mathbb{R}^{4+1} be the five dimensional Minkowski space with the standard Lorentzian metric $m = \text{diag}(-1, 1, 1, 1, 1)$. Denote by $A_\alpha : \mathbb{R}^{4+1} \rightarrow \mathfrak{g}$, $\alpha = 0, \dots, 4$, a connection 1-form taking values in the Lie algebra \mathfrak{g} , and by \mathbf{D}_α the associated covariant differentiation,

$$\mathbf{D}_\alpha B := \partial_\alpha B + [A_\alpha, B],$$

acting on \mathfrak{g} -valued functions B . Introducing the curvature 2-form

$$F_{\alpha\beta} := \partial_\alpha A_\beta - \partial_\beta A_\alpha + [A_\alpha, A_\beta],$$

the *hyperbolic Yang–Mills equation* is the Euler–Lagrange equation associated with the formal Lagrangian action functional

$$\mathcal{L}(A_\alpha) := \frac{1}{2} \int_{\mathbb{R}^{4+1}} \langle F_{\alpha\beta}, F^{\alpha\beta} \rangle dx dt.$$

Here we are using the standard convention of raising indices using the metric m . Thus, the Yang–Mills equation takes the form

$$\mathbf{D}^\alpha F_{\alpha\beta} = 0. \quad (1.1)$$

There is a natural energy-momentum tensor associated to the Yang–Mills equations, namely

$$T_{\alpha\beta} = 2m^{\gamma\delta} \langle F_{\alpha\gamma}, F_{\beta\delta} \rangle - \frac{1}{2} m_{\alpha\beta} \langle F_{\gamma\delta}, F^{\gamma\delta} \rangle.$$

If A solves the Yang–Mills equation (1.1) then $T_{\alpha\beta}$ is divergence-free,

$$\partial^\alpha T_{\alpha\beta} = 0. \quad (1.2)$$

Integrating this for $\beta = 0$ yields a conserved energy

$$\mathcal{E}(A) = \mathcal{E}_{\{t\} \times \mathbb{R}^4}(A) = \int_{\{t\} \times \mathbb{R}^4} T_{00} dx = \sum_{\alpha < \beta} \int_{\{t\} \times \mathbb{R}^4} \langle F_{\alpha\beta}, F_{\alpha\beta} \rangle dx. \quad (1.3)$$

The case $\beta \neq 0$ yields further conservation laws, i.e. the momentum, which play no role in the present work.

The Yang–Mills equation also has a scale invariance property,

$$A(t, x) \rightarrow \lambda A(\lambda t, \lambda x) \quad (\lambda > 0).$$

The energy functional \mathcal{E} is invariant with respect to scaling precisely in dimension $4 + 1$. For this reason we call the $4 + 1$ problem *energy critical*; this is one of the motivations for our interest in this problem.

1.3. Gauge invariance. In order to study the Yang–Mills equation as a well-defined evolution in time one needs to also consider its gauge invariance. Given a map $O = O(t, x)$ taking values in the group \mathbf{G} , we introduce

$$O_{;\alpha} = \partial_\alpha O O^{-1},$$

which now takes values in the Lie algebra \mathfrak{g} . The gauge transformation of a connection A by O is

$$A_\alpha \longrightarrow \text{Ad}(O)A_\alpha - O_{;\alpha} =: \mathcal{G}(O)A_\alpha,$$

which makes the associated differentiation \mathbf{D} covariant with $\text{Ad}(O)$. Correspondingly, the curvature tensor changes by

$$F_{\alpha\beta} \longrightarrow \text{Ad}(O)F_{\alpha\beta}.$$

Clearly, the Yang–Mills equation (1.1) is invariant under such transforms.

As a consequence, solutions are a-priori defined as equivalence classes. In order to uniquely select representatives for the solutions to the Yang–Mills equation within each equivalence class one needs to add an additional set of constraint equations; this procedure is known as *gauge fixing*. This issue is fundamental for the fine analysis of the Yang–Mills equation. In choosing a gauge, one is naturally led to pursue conflicting goals:

- (i) Causality: the system should have finite speed of propagation
- (ii) Structure: the nonlinearity should exhibit null structure type cancellation
- (iii) Large data: the gauge should be well-defined for large data.

Historically there are (at least) three gauges that have played a role in the study of the hyperbolic Yang–Mills evolution:

1. *The Lorenz gauge,*

$$\partial^\alpha A_\alpha = 0.$$

In this gauge the Yang–Mills equation becomes a system of semilinear wave equations for A_α , and in particular it has finite speed of propagation. This gauge is very convenient for local well-posedness for large but regular data. However, it is not so good in the low regularity setting as it does not capture well the null structure, see e.g. [30].

2. *The temporal gauge,*

$$A_0 = 0.$$

This again insures that the above system is strictly hyperbolic, and in particular it has finite speed of propagation. In this gauge the equations can be understood as a semilinear wave equation for the curl of A_x , coupled with a transport equation for its divergence. This gauge is also very convenient for local well-posedness for large but regular data, and it fully describes all regular solutions to the hyperbolic Yang–Mills equation. Again there are multiple technical difficulties if one tries to implement such a gauge in the low regularity setting or globally in time. In particular we have no dispersion for the divergence of A . This gauge will play an auxiliary role in our analysis, and is described in greater detail in Section 4.

3. *The Coulomb gauge,*

$$\sum_{j=1}^4 \partial_j A_j = 0.$$

Here the causality is lost; however, the Coulomb gauge is an “elliptic” gauge which captures well the null structure of the problem, and thus works well in low regularity settings. Indeed, the Coulomb gauge was used in [14] to prove the small data result for this problem. Unfortunately, it seems that the Coulomb gauge cannot be implemented globally for large data, even after restricting to those below the ground state energy. Nevertheless, for expository purposes we do provide a brief review of the Coulomb gauge in the beginning of Section 2.

For the reasons described above, these three gauges seem inadequate for the purpose of proving the Threshold Theorem (to be described below). Instead, in our first article [23] we introduce a new gauge, namely

4. *The caloric gauge.* This is defined via the Yang–Mills heat flow and is described in Section 2. It has the key property that it is globally defined for all data below the ground

state energy. In addition, to the leading order this agrees with the Coulomb gauge, so there are many similarities between the analysis in the caloric and Coulomb gauges.

1.4. Yang–Mills initial data sets. In order to consider the hyperbolic Yang–Mills problem as an evolution equation we need to consider initial data sets. An *initial data set* for (1.1) consists of a pair of 1-forms (a_j, e_j) on \mathbb{R}^4 . We say that (a_j, e_j) is the initial data for a Yang–Mills solution A if

$$(A_j, F_{0j})|_{\{t=0\}} = (a_j, e_j).$$

The curvature of a is denoted by f in what follows.

Note that (1.1) imposes the condition that the following equation be true for any initial data for (1.1):

$$\mathbf{D}^j e_j = 0. \quad (1.4)$$

where \mathbf{D}^j denotes the covariant derivative with respect to the a_j connection. This equation is the *Gauss* (or the *constraint*) *equation* for (1.1).

Definition 1.1. (1) A regular initial data set for the Yang–Mills equation is a pair of 1-forms $(a_j, e_j) \in H_{loc}^N \times H^{N-1}$, $N \geq 2$, with $f \in H^N$, and which satisfies the constraint equation (1.4).

(2) A finite energy initial data set for the Yang–Mills equation is a pair of 1-forms $(a_j, e_j) \in H_{loc}^1 \times L^2$ with $f \in L^2$ and which satisfies the constraint equation (1.4).

1.5. Yang–Mills solutions. We begin by defining the notions of regular and finite energy solutions:

Definition 1.2. (1) Let $N \geq 2$. A regular solution for the Yang–Mills equation in an open set $\mathcal{O} \subset \mathbb{R}^{4+1}$ is a connection $A \in C([0, T]; H_{loc}^N)$, whose curvature satisfies $F \in C([0, T]; H_{loc}^{N-1})$ and which solves the equation (1.1).

(2) A finite energy solution for the Yang–Mills equation in the open set \mathcal{O} is a connection $A \in C([0, T]; H_{loc}^1)$, whose curvature satisfies $F \in C([0, T]; L_{loc}^2)$ and which is the limit of regular solutions in this topology.

We carefully remark that this definition does not require a gauge choice. Hence at this point solutions are still given by equivalence classes. Corresponding to the above classes of solutions, we have the classes of gauge transformations which preserve them:

Definition 1.3. (1) Let $N \geq 2$. A regular gauge transformation in an open set $\mathcal{O} \subset \mathbb{R}^{4+1}$ is a map

$$O : \mathcal{O} \rightarrow \mathbf{G}$$

with the following regularity properties:

$$O_{;x}, O_{;t} \in C_t(H_{loc}^{N+1}).$$

(2) An admissible gauge transformation in an open set $\mathcal{O} \subset \mathbb{R}^{4+1}$ is a similar map with the following regularity properties:

$$O_{;x}, O_{;t} \in C_t(H_{loc}^1).$$

Using this notion we can now talk about gauge equivalent connections:

Definition 1.4. Two finite energy connections $A^{(1)}$ and $A^{(2)}$ in an open set $\mathcal{O} \subset \mathbb{R}^{4+1}$ are gauge equivalent if there exists an admissible gauge transformation O so that $A^{(2)} = OA^{(1)}O^{-1} - O_{;x}$.

1.6. Topological classes. The space of finite energy Yang–Mills connections in \mathbb{R}^4 is not connected. Instead, such connections can be classified in terms of their *topological class*; see Section 4 for more details.

For a compact base manifold, such as \mathbb{S}^4 , this term refers to the isomorphism classes of principal \mathbf{G} -bundles which supports the connection. On the other hand, for \mathbb{R}^4 , which is contractible and thus supports only the trivial fiber bundles, a topological class must be interpreted rather as a property of a connection.

In the particular case of four dimensional $SU(2)$ connections the topological class is easily described in terms of the (second) Chern number

$$c_2 = \frac{1}{8\pi^2} \int_{\mathbb{R}^4} \text{tr}(F \wedge F).$$

This is always an integer if A has finite energy. For an arbitrary compact noncommutative Lie group, we have an analogue of c_2 ,

$$\chi(A) = \int_{\mathbb{R}^4} -\langle F \wedge F \rangle = \frac{1}{4} \int_{\mathbb{R}^4} -\langle F_{ij}, F_{kl} \rangle dx^i \wedge dx^j \wedge dx^k \wedge dx^\ell,$$

which we denote by $\chi(A)$ and call the *characteristic number*. This quantity is still a topological invariant, but it no longer fully describes the topological class.

The connections which are in the same class as the zero connection are called *topologically trivial*. For such connections, $\chi = 0$. An alternative way to describe topologically trivial connections is given by the following result, which generalizes Uhlenbeck’s lemma [44]:

Theorem 1.5 ([25]). *A finite energy connection A in \mathbb{R}^4 is topologically trivial iff $A \in \dot{H}^1$ in a suitable gauge.*

A further “Good Global Gauge Theorem” is provided in [25] for finite energy connections which are not topologically trivial.

1.7. Solitons and the ground state energy. Steady states for the hyperbolic Yang–Mills equation are called *harmonic Yang–Mills connections*, and play an important role in our work. They solve the equations

$$\mathbf{D}^j F_{kj} = 0 \quad \text{in } \mathbb{R}^4, \tag{1.5}$$

and can be seen as critical points for the Lagrangian

$$\mathcal{E}_e(A) = \frac{1}{2} \int_{\mathbb{R}^{4+1}} \langle F_{ij}, F^{ij} \rangle dx dt.$$

The key elliptic regularity result is as follows:

Theorem 1.6 (Uhlenbeck [44, 45]). *\dot{H}^1 harmonic Yang–Mills connections are smooth in a suitable gauge.*

The question of existence of finite energy harmonic Yang–Mills connections is best phrased in terms of the topological classes described above:

Theorem 1.7. *The following properties hold for harmonic Yang–Mills connections:*

- (1) *Within each topological class there exist energy minimizers. These are called instantons, and come in two varieties, self-dual $F = \star F$ and anti-self-dual $F = -\star F$, depending on the topological class.*

(2) In particular, there exists a unique (up to symmetries) minimal energy nontrivial harmonic Yang–Mills connection Q , which is necessarily an instanton, whose energy E_{GS} satisfies

$$\mathcal{E}(Q_{GS}) = |\chi(Q_{GS})|.$$

(3) All nontrivial harmonic Yang–Mills connection a , with energy $\mathcal{E}(Q) \leq 2E_{GS}$ are instantons and satisfy

$$\mathcal{E}(Q) = |\chi(Q)|.$$

Parts (1) & (2) are classical. We remark that part (3), which follows from a recent result of [4], is nontrivial due to existence of non-minimizing harmonic Yang–Mills connections [31]. We refer to [25, Sections 1.8 and 6] for further discussion.

As a consequence of the above properties, it easily follows that in the class of topologically trivial connections, the threshold for nontrivial harmonic Yang–Mills connections is $2E_{GS}$ rather than E_{GS} .

We also remark that harmonic Yang–Mills connections which are not energy minimizers no longer have to be self-dual or anti-self-dual.

The harmonic Yang–Mills connections are relevant for the hyperbolic Yang–Mills flow for multiple reasons. First of all, they provide examples of solutions that do not scatter. Further, above the ground state energy E_{GS} there are examples of solutions which blow up in finite time, with a profile which approaches a rescaled instanton, see [9, 27]. Thus, the ground state energy arises as a natural threshold in the large data well-posedness theory, and one is led to the *Threshold Conjecture*, which asserts that the Yang–Mills problem is globally well-posed below the ground state energy. All such connections must be topologically trivial. However, as discussed above, for such connections the correct threshold is $2E_{GS}$. Based on the above discussion, we will call *subthreshold data/solution* any topologically trivial hyperbolic Yang–Mills data/solution with energy below $2E_{GS}$.

1.8. The main results. The main question we are concerned with is whether the hyperbolic Yang–Mills equation (1.1) is globally well-posed in the space of finite energy connections in the $4 + 1$ dimensional setting. The small data global well-posedness was recently proved by Krieger together with the second author in [14], so our main interest here is in large solutions. The Threshold Conjecture asserts that global well-posedness in the energy space holds below the ground state energy.

The first goal of our four papers [23, 24, 25, 26] is to establish the validity of (a more precise form of) this conjecture. In the simplest form, our result can be phrased as follows:

Theorem 1.8 (Threshold Theorem for Energy Critical Yang–Mills). *Global well-posedness and scattering holds for the energy critical hyperbolic Yang–Mills evolution in \mathbb{R}^{4+1} for all topologically trivial initial data with energy below $2E_{GS}$.*

Since scattering solutions are necessarily topologically trivial, we are justified in considering only the topologically trivial data in Theorem 1.8. This restriction, in view of Theorem 1.7, is the reason why our threshold is $2E_{GS}$ rather than just E_{GS} .

The statement of this theorem should be understood as follows:

- For each smooth subthreshold initial data (a, e) there exists a global smooth solution, which is unique up to gauge transformations.

- For each subthreshold data in $\dot{H}^1 \times L^2$ there exists a solution $(A, \partial_t A) \in C(\mathbb{R}; \dot{H}^1 \times L^2)$ which is the unique limit of smooth solutions up to gauge transformations.

The above formulation of the result is gauge independent. However, in order to both prove this result and to provide a better description of the solutions, including their scattering properties, it is essential to fix the gauge choice in a favorable way. For our problem, the classical choices of gauge (Lorenz, temporal or Coulomb) seem to present different but equally insurmountable difficulties. We instead rely on the *caloric gauge*, which is constructed based on the regularity theory of the Yang–Mills heat flow, the parabolic counterpart of (1.1). A gauge dependent formulation of this result will be provided later on, see Theorem 5.3.

The second goal of our four papers [23, 24, 25, 26] is to also consider solutions which do not satisfy the topological and energy constraint of the Threshold Theorem. Then on the one hand, we know there exist solutions which blow-up or are global but do not scatter, see [9, 27]. On the other hand scattering can only hold for topologically trivial solutions. Because of this, our second result offers a dichotomy:

Theorem 1.9 (Dichotomy Theorem for Energy Critical Yang–Mills). *The energy critical hyperbolic Yang–Mills evolution in \mathbb{R}^{4+1} is locally well-posed in the energy space. Further, one of the following two properties must hold for the maximal solution:*

- (i) *The solution is topologically trivial, global and scatters at infinity.*
- (ii) *The solution bubbles off a soliton either*
 - (a) *at a finite blow-up time, or*
 - (b) *at infinity.*

We note that these two alternatives hold separately for positive and negative time. In other words we do not eliminate the scenario where, say, scattering holds for positive time while finite time blow-up occurs for negative time.

To fully describe this result we need to clarify the meaning of bubbling off. We do this in the two scenarios, of finite time blow-up solutions and of global solutions.

a) The finite time blow-up scenario: Let $t_0 > 0$ be the blow-up time (maximal existence time) for a finite energy Yang–Mills connection A . By energy conservation, finite speed of propagation and the small data result there must exist a point $x_0 \in \mathbb{R}^4$ so that energy concentrates in the backward blow-up cone centered at (t_0, x_0) , namely $C = \{|x - x_0| < t_0 - t\}$, in the sense that

$$\lim_{t \nearrow t_0} \mathcal{E}_{S_t}(A) > 0.$$

where $S_t = C \cap (\{t\} \times \mathbb{R}^4)$.

In this context, we say that A *bubbles off a soliton at (t_0, x_0)* if there exists a sequence of points $(t_n, x_n) \rightarrow (t_0, x_0)$ and scales r_n with the following properties:

- (1) Time-like concentration,

$$\limsup_{n \rightarrow \infty} \frac{x_n - x_0}{|t_n - t_0|} = v, \quad |v| < 1$$

- (2) Below self-similar scale,

$$\limsup_{n \rightarrow \infty} \frac{r_n}{|t_n - t_0|} = 0$$

(3) Convergence to soliton:

$$\lim_{n \rightarrow \infty} r_n \mathcal{G}(O_n) A(t_n + r_n t, x_n + r_n x) = L_v Q(t, x) \quad \text{in } H_{loc}^1([-1/2, 1/2] \times \mathbb{R}^4)$$

for some sequence of admissible gauge transformations O_n , a Lorentz transformation L_v and finite energy harmonic Yang–Mills connection Q .

We remark that for a finite energy harmonic Yang–Mills connection Q we must have

$$\mathcal{E}(Q) \leq \mathcal{E}(L_v Q)$$

with equality iff $v = 0$.

b) Global solutions. Here we consider a finite energy Yang–Mills connection A which is global forward in time. We say that A *bubbles off a soliton at infinity* if there exists a sequence of points $C \ni (t_n, x_n) \rightarrow \infty$ and scales r_n with the following properties:

(1) Time-like concentration,

$$\limsup_{n \rightarrow \infty} \frac{x_n}{t_n} = v, \quad |v| < 1$$

(2) Below self-similar scale,

$$\limsup_{n \rightarrow \infty} \frac{r_n}{t_n} = 0$$

(3) Convergence to soliton:

$$\lim_{n \rightarrow \infty} r_n \mathcal{G}(O_n) A(t_n + r_n t, x_n + r_n x) = L_v Q(t, x) \quad \text{in } H_{loc}^1([-1/2, 1/2] \times \mathbb{R}^4)$$

for some sequence of admissible gauge transformations O_n , a Lorentz transformation L_v and finite energy harmonic Yang–Mills connection Q .

The proof of these two theorems is the final outcome of the sequence of papers [23], [24], [25] and [26]. These contain conceptually disjoint, self-contained logical steps which address different aspects of the problem, as follows:

I. The caloric gauge [23]: This first paper uses the *Yang–Mills heat flow* in order to introduce the *caloric gauge*, which is central in our analysis. Its main outcome is to provide a complete caloric gauge representation for the hyperbolic Yang–Mills equation (1.1). Along the way, we also establish the Threshold and the Dichotomy Theorems for the Yang–Mills heat flow. In particular, the former allows us to prove that all subthreshold data admit a caloric representation. These results are discussed in Section 2.

II. Energy dispersed solutions [24]: Here we develop the analytic tools which are needed in order to understand the hyperbolic Yang–Mills flow in the caloric gauge. The main result is a strong quantitative a-priori bound for *energy dispersed solutions*, which in particular implies local well-posedness as well as small data global well-posedness in the caloric gauge. The notion of energy dispersion as well as the main results are described in Section 3.

III. Large data and causality [25]: Since not all Yang–Mills solutions can be placed in the caloric gauge, in this article we show how to switch the qualitative part of the analysis (but not the analytic part) into the temporal gauge, in order to be able to deal with data with above threshold energy. The overview in Section 4 also covers

topological classes, initial data surgery and gauge matters such as patching of local solutions.

IV. Blow-up analysis [26]: In this final step we use Morawetz type bounds in order to perform a blow-up analysis which leads to the proof of the two theorems above. This is where the results in the previous two papers [24] and [26] are used, but not the analysis leading to these results. This is described in the last section.

We finally remark that these papers build upon a large body of work. This begins with early results on Yang–Mills above scaling [17, 2, 3, 6, 8], where the structure of the equations was first understood and exploited. Our general approach broadly follows the outline of similar results for wave maps, starting with the small data problem, the null frame function spaces and the renormalization idea [42, 35, 43] and continuing with the induction on energy based energy dispersion approach in the proof of the Threshold and Dichotomy Theorem in [32, 33] (see also [11] and [39, 37, 38, 40, 41]). The similar results for the closely related Maxwell–Klein–Gordon equation at critical regularity were proved in the small data case in [28] ($d \geq 6$) and [13] ($d \geq 4$), respectively large data in [21, 22, 20] and independently in [10]. Finally, the small data results for (YM) were obtained only recently in [12] ($d \geq 6$) and [14] ($d \geq 4$). For a more extensive overview of related literature we refer the reader to [26]. Some further comments are provided in each of the following sections as needed.

Acknowledgments. Part of the work was carried out during the semester program “New Challenges in PDE” held at MSRI in Fall 2015. S.-J. Oh was supported by the Miller Research Fellowship from the Miller Institute, UC Berkeley and the TJ Park Science Fellowship from the POSCO TJ Park Foundation. D. Tataru was partially supported by the NSF grant DMS-1266182 as well as by the Simons Investigator grant from the Simons Foundation.

2. THE CALORIC GAUGE

This section describes the main results of [23], whose aim is to develop the *caloric gauge* as our main gauge of choice in the study of the hyperbolic Yang–Mills evolution.

Let us take as a starting point of our discussion the following small data result proved earlier in [14]:

Theorem 2.1. *The hyperbolic Yang–Mills equation in \mathbb{R}^{4+1} is globally well-posed in the Coulomb gauge for all initial data with small energy.*

Unfortunately, while the Coulomb gauge works well in the small data problem, it does not appear to work for large data, even after restricting to only subthreshold data. This large data difficulty with the Coulomb gauge compels us to look for a different gauge choice, in which the Yang–Mills equation exhibits a similar null structure as the Coulomb gauge, yet which can be used in the large data problem.

Our solution to this problem is to introduce and use the (*global*) *caloric gauge*, which is constructed with the help of the *Yang–Mills heat flow*. A more localized form of this gauge was previously introduced by the first author in [18, 19], in order to study local well-posedness questions for the $3 + 1$ dimensional hyperbolic Yang–Mills equation. This was in turn inspired by Tao’s caloric gauge for wave maps [36], which is based on the harmonic map heat flow.

On the one hand, the caloric gauge resembles Coulomb gauge in the sense that a generalized Coulomb condition holds (to be discussed in more detail in Section 2.4). On the other hand, it can be used for a larger class of connections, which in particular includes all subthreshold connections (essentially by the Threshold Theorem for the Yang–Mills heat flow, see Theorem 2.4 below). Therefore, it furnishes a natural setting to state and prove the Threshold Theorem for the hyperbolic Yang–Mills equation; see Theorem 5.3 below.

2.1. The Coulomb gauge and the null structure. Before we describe the caloric gauge, we first review the null structure of the hyperbolic Yang–Mills equation in the Coulomb gauge, which plays essential role in low regularity problems for the Yang–Mills equation.

Consider the expansion of the Yang–Mills equation (1.1) in terms of A , which takes the form

$$\square A_\beta + 2[A_\alpha, \partial^\alpha A_\beta] = \partial_\beta \partial^\alpha A_\alpha - [\partial^\alpha A_\alpha, A_\beta] + [A^\alpha, \partial_\beta A_\alpha] - [A^\alpha, [A_\alpha, A_\beta]]. \quad (2.1)$$

where $\square_A := \mathbf{D}^\alpha \mathbf{D}_\alpha$ is the covariant d'Alembertian (or the covariant wave operator). Separating the spatial part and the temporal part of the connection, one immediately sees that the spatial divergence of the solutions plays a prominent role. Precisely, one can rewrite the equations in the form

$$\begin{aligned} \square_A A_j &= \partial_j \partial^k A_k + \partial_j \partial^0 A_0 + [A^\alpha, \partial_j A_\alpha] \\ \Delta_A A_0 &= \partial_0 \partial^j A_j + [A^j, \partial_0 A_j]. \end{aligned} \quad (2.2)$$

Thus, when imposing the Coulomb gauge condition,

$$\sum_{j=1}^4 \partial_j A_j = 0, \quad (2.3)$$

the above equations turn into a hyperbolic system for the main variables

$$\square_A A_j = \partial_j \partial^0 A_0 + [A^\alpha, \partial_j A_\alpha].$$

In order to eliminate the first term on the right and also to restrict the evolution to divergence free fields A_j we apply the Leray projection \mathbf{P} , and rewrite the equation in the form

$$\square A_j = \mathbf{P} ([A^\alpha, \partial_j A_\alpha] - 2[A^\alpha, \partial_\alpha A_j] - [\partial_0 A_0, A_j] - [A^\alpha, [A_\alpha, A_j]]). \quad (2.4)$$

Here the A_0 component plays an auxiliary role, and is determined at each fixed time via the elliptic equation

$$\Delta_A A_0 = [A^j, \partial_0 A_j]. \quad (2.5)$$

This does not yet yield a self-contained system, as the time derivative of A_0 also appears in the first equation. A slightly more involved computation yields the equation

$$\partial^j \mathbf{D}_j \mathbf{D}^0 A_0 = \partial^j (2[A_0, \partial^0 A_j] + [\partial_j A_\alpha, A^\alpha] + [A_\alpha, [A^\alpha, A_j]]) \quad (2.6)$$

which serves to also determine $\mathbf{D}^0 A_0$ in an elliptic fashion.

As one can easily see above, the Yang–Mills equations in the Coulomb gauge can be viewed as an evolution equation (2.4) for the spatial part A_x of the connection, whereas A_0 and $\mathbf{D}^0 A_0$ play the role of auxiliary, dependent variables. All terms in the equation which involve A_0 can be thought of as having more of an elliptic character, and to a large extent have a perturbative nature. The quadratic terms

$$\mathbf{P} ([A^k, \partial_j A_k] - 2[A^k, \partial_k A_j])$$

can be thought of as the leading part of the nonlinearity. It is crucial that these terms satisfy the cancellation property known as the null condition.

As mentioned before, the Coulomb gauge works well for the small data problem (Theorem 2.1). Concerning large data, however, one sees here that in order to properly set up the Yang–Mills equation in the Coulomb gauge one would need to be able to invert the operator $\partial^j \mathbf{D}_j$. Exactly the same operator arises when one considers the linearization of the Coulomb gauge condition. This works well in the small data problem, but not so well for the large data problem.

2.2. Local and global theory for the Yang–Mills heat flow. Neglecting for the moment the time component of the connection A , at fixed time we consider the energy functional

$$\mathcal{E}_e(A_x) = \frac{1}{2} \int_{\mathbb{R}^4} \langle F_{ij}, F^{ij} \rangle dx.$$

The *Yang–Mills heat flow* is the gradient flow associated to this functional, which has the expression

$$\partial_s A_i = \mathbf{D}^\ell F_{\ell i}, \quad A_i(s=0) = a_i. \quad (2.7)$$

As written this system is invariant with respect to purely spatial gauge transforms. To better frame the discussion, we observe that one can add a heat time component to the connection A and rewrite the Yang–Mills heat flow in a fully covariant fashion as

$$F_{si} = \mathbf{D}^\ell F_{\ell i}. \quad (2.8)$$

Then one can view the Yang–Mills heat flow equations in (2.7) as the effect of a gauge choice

$$A_s = 0,$$

(which we call the *local caloric gauge*) applied to the fully covariant Yang–Mills heat flow. This is akin to using the temporal gauge for the hyperbolic Yang–Mills equation.

We start with the basic result:

Theorem 2.2. *The problem (2.7) is locally well-posed for data $a \in \dot{H}^1$.*

The assumption $a \in \dot{H}^1$ restricts a (and thus the solution) to the topologically trivial class. This is natural in view of our goal of constructing the caloric gauge, and also for the eventual application to the Threshold Theorem (Theorem 1.8).

In the study of (2.7), a key role is played by the $L^3_{s,x}$ norm of the curvature F_{ij} . Precisely, the solution to (2.7) can be continued and uniform covariant parabolic estimates for the solution can be proved for as long as $\|F\|_{L^3}$ remains finite. This motivates the following definition for the *caloric size* of a connection a :

$$\mathcal{Q}(a) = \begin{cases} \int_{R^+ \times \mathbb{R}^4} |F(s, x)|^3 ds dx & \text{if the solution to (2.7) is global,} \\ \infty & \text{otherwise.} \end{cases}$$

We note that this is a scaling- and gauge-invariant quantity.

As described below, the caloric gauge is defined only for connections a for which $\mathcal{Q}(a)$ is finite. This is an open subset of \dot{H}^1 , as $\mathcal{Q}(a)$ has a locally Lipschitz dependence on a whenever finite. Furthermore, for such a we can describe the behavior of its Yang–Mills heat flow at infinity as follows:

Theorem 2.3 ([23]). *Let $a \in \dot{H}^1$ be a connection so that $\mathcal{Q}(a) < \infty$. Then the corresponding solution has the property that the limit*

$$\lim_{s \rightarrow \infty} A(s) = a_\infty$$

exists in \dot{H}^1 . Further, the limiting connection is flat, $f_\infty = 0$.

The main technical difficulty with (2.7) is that it is only degenerate parabolic. Precisely, (2.7) can be formally viewed as a coupling of a strongly parabolic system for F (which we think of as the curl of A) and a transport equation for the divergence of A .

We note that there is an alternate gauge choice which circumvents this issue, namely the *de Turck gauge*

$$A_0 = \partial^j A_j,$$

where the Yang–Mills heat flow becomes strongly parabolic and is easier to solve locally. In our formalism, the classical de Turck trick of compensating the degeneracy by a suitable s -dependent gauge transformation amounts to solving (2.8) in this gauge, hence the name.

Unfortunately, the transition from local to global is impossible in the de Turck gauge; in other words, Theorem 2.3 is false in the de Turck gauge. One can see this by considering the evolution of flat connections. This is trivial under the local caloric gauge, but yields a $4 + 1$ dimensional harmonic heat flow for maps into \mathbf{G} in the de Turck gauge, which is known to possibly blow up.

Our approach is instead based on a version of the de Turck trick for the *linearization* of (2.7) (namely, (2.12) below). In this scheme, an auxiliary flow called the *dynamic Yang–Mills heat flow* plays a major role. We will return to discussion of this idea in Section 2.6.

For now, we proceed to describe our next result proved in [23], which asserts that all connections with energy below threshold $2E_{GS}$ have finite caloric size, and thus Theorem 2.3 applies:

Theorem 2.4 (Threshold Theorem for the heat flow). *There exists a nondecreasing function*

$$\mathcal{Q} : [0, 2E_{GS}) \rightarrow [0, \infty)$$

so that for every connection 1-form $a \in \dot{H}^1$ with subthreshold energy $\mathcal{E} < 2E_{GS}$, we have

$$\mathcal{Q}(a) \leq \mathcal{Q}(\mathcal{E}) \tag{2.9}$$

This is proved using a concentration compactness type argument. The key ingredient is the *energy monotonicity formula*

$$\mathcal{E}_e(A(s_1)) - \mathcal{E}_e(A(s_2)) = - \int_{s_1}^{s_2} \int \langle \mathbf{D}^\ell F_{\ell j}, \mathbf{D}^k F_k{}^j \rangle dx ds.$$

This formula yields good control of A in the local caloric gauge, but not in the de Turck gauge. The same argument also gives the corresponding Dichotomy Theorem:

Theorem 2.5 (Dichotomy Theorem for the heat flow). *For any $a \in \dot{H}^1$, one of the following two properties must hold for the maximally extended solution:*

- (i) *The solution is topologically trivial, global and $\mathcal{Q}(a) < \infty$.*
- (ii) *The solution bubbles off a harmonic Yang–Mills connection either*
 - (a) *at a finite blow-up time, or*
 - (b) *at infinity.*

The bubbling argument here has roots in the classical work of Struwe [34] (see also Schlatter [29]) on compact manifolds. In comparison, the significance of the above theorems lies in the precise asymptotics of the Yang–Mills heat flow on the noncompact space \mathbb{R}^4 , which allows us to construct the caloric gauge.

2.3. Caloric connections and the caloric manifold. Since the limiting connection a_∞ given by Theorem 2.3 is flat, it must be gauge equivalent to the zero connection. Precisely, there exists a gauge transformation O with the property that

$$a_{\infty,j} = O^{-1}\partial_j O.$$

Here $O = O(a) \in \dot{H}^2$ (interpreted in the sense that $O_{;j} := \partial_j O O^{-1} \in \dot{H}^1$) is unique up to constant gauge transformations. Conjugating the full heat flow with respect to such an O yields a gauge equivalent connection

$$\tilde{A}_j = O A_j O^{-1} - O_{;j},$$

which solves the Yang–Mills heat flow, and satisfies $\tilde{a}_\infty = 0$. This lead us to the following definition of *caloric* connections:

Definition 2.6. We will say that a connection $a \in \dot{H}^1$ is caloric if $a_\infty = 0$. We denote the set of all such connections by \mathcal{C} .

Theorem 2.4 can then be restated as an existence result for gauge equivalent caloric connections:

Theorem 2.7 ([23]). *For every connection $a \in \dot{H}^1$ with $\mathcal{Q}(a) < \infty$ there exists a gauge equivalent caloric connection $\tilde{a} \in \dot{H}^1$, which is unique up to constant gauge transformations. In particular, this conclusion holds for all subthreshold connections.*

The connection \tilde{a} is defined as

$$\tilde{a}_j = O a_j O^{-1} - O_{;j}, \quad O = O(a).$$

We note that the two connections have the same caloric size, $\mathcal{Q}(a) = \mathcal{Q}(\tilde{a})$.

To solve the Yang–Mills equation in the caloric gauge we need to view the family \mathcal{C} of the caloric gauge connections with energy below the ground state energy as an infinite dimensional manifold. Here the \dot{H}^1 topology is no longer sufficient, so we introduce the slightly stronger topology

$$\mathbf{H} = \{a \in \dot{H}^1 : \partial^j a_j \in \ell^1 L^2\}$$

which reflects the fact, to be discussed later in more detail, that caloric connections satisfy a generalized, nonlinear form of the Coulomb gauge condition. Then we have

Theorem 2.8 ([23]). *For any caloric subthreshold connections a with energy \mathcal{E} and caloric size \mathcal{Q} we have the \mathbf{H} bound*

$$\|a\|_{\mathbf{H}} \lesssim_{\mathcal{E}, \mathcal{Q}} 1 \tag{2.10}$$

The set \mathcal{C} of all \dot{H}^1 caloric connections is a C^1 infinite dimensional submanifold of \mathbf{H} .

We denote

$$\tilde{a} = \text{Cal}(a).$$

For arbitrary subthreshold $a \in \dot{H}^1$ this is only defined as an equivalence class, modulo constant conjugations. However, if in addition we know that $a \in \mathbf{H}$, then $O(a)$ is continuous, and we can fix its choice by imposing the additional condition

$$\lim_{x \rightarrow \infty} O(x) = Id. \quad (2.11)$$

With this choice we have the following regularity property:

Theorem 2.9. *The map $a \rightarrow O(a)$ is continuous (though not Lipschitz) from \dot{H}^1 to \dot{H}^2 . It is also locally C^1 from \mathbf{H} to $\dot{H}^2 \cap C^0$.*

2.4. The tangent space and caloric data sets. Finite energy caloric Yang–Mills waves will be continuous functions of time which take values into \mathcal{C} . They are however not smooth in time, instead their time derivative will merely belong to L^2 . Because of this, we need to take the closure of its tangent space $T\mathcal{C}$ (which a-priori is a closed subspace of \mathbf{H}) in L^2 . This is denoted by $T_a^{L^2}\mathcal{C}$. It is also convenient to have a direct way of characterizing this space; that is naturally done via the linearization of the caloric flow:

Definition 2.10. For a caloric gauge connection $a \in \mathcal{C}$, we say that $L^2 \ni b \in T_a^{L^2}\mathcal{C}$ iff the solution to the linearized local caloric gauge Yang–Mills heat flow equation

$$\partial_s B_k = [B^j, F_{kj}] + \mathbf{D}^j(\mathbf{D}_k B_j - \mathbf{D}_j B_k), \quad B_k(0) = b_k \quad (2.12)$$

satisfies

$$\lim_{s \rightarrow \infty} B(s) = 0.$$

Turning our attention now to the Yang–Mills flow, we will now consider solutions which at any fixed time t are in the caloric gauge, $A_x(t) \in \mathcal{C}$.

Definition 2.11. An initial data for the Yang–Mills equation in the caloric gauge is a pair (a, b) where $a \in \mathcal{C}$ and $b_k \in T_a^{L^2}\mathcal{C}$.

The transition from one time to another requires understanding the linearization of the Yang–Mills heat flow. As in the Coulomb gauge, we will consider the spatial component of the connection as the dynamic variable, and view the temporal part of the connection as an auxiliary variable. We begin our discussion by considering the initial data. To connect a general initial data (a_k, e_k) with caloric initial data we have the following result:

Theorem 2.12. (1) *For any initial data pair $(a, e) \in \dot{H}^1 \times L^2$ with finite caloric size, there exists a caloric gauge data set $(\tilde{a}, b) \in T^{L^2}\mathcal{C}$ and $a_0 \in \dot{H}^1$, unique up to constant gauge transformations and with continuous dependence in this quotient topology, so that (\tilde{a}, \tilde{e}) is gauge equivalent to (a, e) and*

$$\tilde{e}_k = b_k - (\mathbf{D}_{\tilde{a}})_k a_0.$$

(2) *For any caloric gauge initial data set $(\tilde{a}, b) \in T^{L^2}\mathcal{C}$, there exists a unique $a_0 \in \dot{H}^1$, with Lipschitz dependence on $(a, b) \in \dot{H}^1 \times L^2$, so that*

$$e_k = b_k - (\mathbf{D}_a)_k a_0$$

satisfies the constraint equation (1.4).

¹Here \dot{H}^2 needs to be interpreted as a quotient space, modulo constant conjugations

²Here the action of the group of constant conjugations can be eliminated by using the condition (2.11).

In view of this result, we can fully describe caloric Yang–Mills waves as continuous functions

$$I \ni t \rightarrow (A_x(t), \partial_0 A_x(t)) \in T^{L^2} \mathcal{C}.$$

An important role in the proof of this theorem is played by the following nonlinear div-curl type decomposition for the tangent space $T_a^{L^2} \mathcal{C}$:

Theorem 2.13. *Let $a \in \mathcal{C}$ with energy \mathcal{E} and caloric size \mathcal{Q} . Then for each $e \in L^2$ there exists a unique decomposition*

$$e = b - \mathbf{D}a_0, \quad b \in T_a^{L^2} \mathcal{C}, \quad a_0 \in \dot{H}^1. \quad (2.13)$$

with the corresponding bound

$$\|b\|_{L^2} + \|a_0\|_{\dot{H}^1} \lesssim_{\mathcal{E}, \mathcal{Q}} \|e\|_{L^2}. \quad (2.14)$$

Proving the latter theorem, in turn, requires understanding of the linearized equation (2.12); we will return to this issue in Section 2.6.

2.5. The dynamic Yang–Mills heat flow and the hyperbolic Yang–Mills equation.

To proceed further, given a caloric Yang–Mills wave on I , we seek to interpret the (covariant) hyperbolic Yang–Mills equation

$$\mathbf{D}^\alpha F_{\alpha\beta} = 0, \quad (2.15)$$

as gauge dependent hyperbolic evolutions for A_x . Separating these equations into

$$\mathbf{D}^\alpha \mathbf{D}_\alpha A_k = \mathbf{D}^k \mathbf{D}^\alpha A_\alpha - [A^\alpha, \mathbf{D}_k A_\alpha], \quad (2.16)$$

respectively,

$$\mathbf{D}^k \mathbf{D}_k A_0 = \mathbf{D}_0 \mathbf{D}^k A_k - [A^k, \mathbf{D}_0 A_k], \quad (2.17)$$

we seek to interpret the first equation as a hyperbolic evolution for A_x , and the second as an elliptic compatibility condition for A_0 . This is achieved in several steps as follows:

(i) First, we show that the pair $(A_x, \partial_0 A_x) \in T^{L^2} \mathcal{C}$ satisfies a generalized Coulomb like condition,

$$\partial^k A_k = \mathbf{D}\mathbf{A}(A), \quad \partial^k A_k = \mathbf{D}\mathbf{B}(A, B), \quad (2.18)$$

where $\mathbf{D}\mathbf{A}$ and $\mathbf{D}\mathbf{B}$ are nice maps on $T^{L^2} \mathcal{C}$, which contains an explicitly computed quadratic part, as well as purely perturbative higher order terms. Of course, this step does not have to do with (2.15), and holds for any pair in $T^{L^2} \mathcal{C}$. The key computation for $\partial^k A_k$ is

$$\partial^k A_k = - \int_0^\infty \partial^k \partial_s A_k(s) ds = - \int_0^\infty \mathbf{D}^k F_{sk}(s) + (\text{quadratic and higher})$$

but by (2.7), the linear term vanishes. A similar computation holds for $\partial^k B_k$.

(ii) Next, we use the $\beta = 0$ part of the equation (2.15) to show that A_0 is uniquely determined by A_x and $B_x = \partial_0 A_x$, i.e.,

$$A_0 = \mathbf{A}_0(A_x, B_x)$$

where \mathbf{A}_0 is a nice smooth map on $T^{L^2} \mathcal{C}$ which contains an explicitly computed quadratic part, as well as purely perturbative higher order terms.

(iii) Moreover, we use the $\beta \neq 0$ part of the equation (2.15) to show that $\mathbf{D}^0 A_0$ is uniquely determined by A_x and $B_x = \partial_0 A_x$,

$$\mathbf{D}^0 A_0 = \mathbf{D}\mathbf{A}_0(A_x, B_x)$$

where $\mathbf{D}\mathbf{A}_0$ is a nice smooth map on $T^{L^2}\mathcal{C}$ which again contains an explicitly computed quadratic part, as well as purely perturbative higher order terms.

The above steps allow us, just as in the case of the Coulomb gauge, to view the spatial part of the connection $(A_x, \partial_0 A_x) \in T^{L^2}\mathcal{C}$ as the dynamical variable, and $A_0, \partial_0 A_0$ as dependent variables. Precisely, we can recast the equations (2.16) in the form

$$\square_A A_k = \mathbf{P}[A_x, \partial_k A_x] + 2\Delta^{-1} \partial_k \mathbf{Q}(\partial^\alpha A_x, \partial_\alpha A_x) + R(A, \partial_t A), \quad (2.19)$$

where $[A_x, B_x]$ is a shorthand for $[A^\ell, B_\ell]$, and \mathbf{Q} is a symmetric bilinear form with symbol³

$$\mathbf{Q}(\xi, \eta) = \frac{\xi^2 - \eta^2}{2(\xi^2 + \eta^2)}.$$

Here on the right we have two explicit quadratic terms depending only on A_x , and its time derivative, both of which have a favorable null structure, and a remainder higher order term R which admits favorable $L^1 L^2$ bounds and thus only plays a perturbative role. However, in the covariant d'Alembertian on the left, we still have the coefficients A_0 and $\mathbf{D}_0 A_0$, which are determined as above in terms of A_x and $\partial_t A_x$:

$$\begin{aligned} A_0 &= \mathbf{A}_0(A_x, B_x) = \mathbf{A}_0^2(A_x, B_x) + \mathbf{A}_0^3(A_x, B_x), \\ \mathbf{D}_0 A_0 &= \mathbf{D}\mathbf{A}_0(A_x, B_x) = \mathbf{D}\mathbf{A}_0^2(B_x, B_x) + \mathbf{D}\mathbf{A}_0^3(A_x, B_x), \end{aligned} \quad (2.20)$$

Here the quadratic terms $\mathbf{A}_0^2(A_x, B_x), \mathbf{D}\mathbf{A}_0^2(A_x, B_x)$ are explicit translation invariant bilinear forms,

$$\mathbf{A}_0^2(A_x, B_x) = \Delta^{-1}[A_x, B_x] + 2\Delta^{-1} \mathbf{Q}(A_x, B_x), \quad (2.21)$$

$$\mathbf{D}\mathbf{A}_0^2(B_x, B_x) = -2\Delta^{-1} \mathbf{Q}(B_x, B_x). \quad (2.22)$$

The remainders $\mathbf{A}_0^3(A_x, B_x), \mathbf{D}\mathbf{A}_0^3(A_x, B_x)$, however, are not explicit but satisfy favorable bounds. Of these only the quadratic part of A_0 plays a nonperturbative role.

Finally, A_x is also subject to a compatibility condition

$$\partial^k A_k = \mathbf{D}\mathbf{A}(A) := \mathbf{Q}(A, A) + \mathbf{D}\mathbf{A}^3(A), \quad (2.23)$$

where $\mathbf{D}\mathbf{A}^3$ is perturbative.

To study the small data problem it would be sufficient to work with the equation (2.19). However, for the large data problem we also need to flow the wave equation in the parabolic direction, which in turn requires us to specify the s -evolution equation for A_0 . Our choice is to use the *dynamic Yang–Mills heat flow*

$$F_{s\alpha} = \mathbf{D}^\ell F_{\ell\alpha}, \quad (2.24)$$

which is the (covariant) Yang–Mills heat flow (2.8) adjoined with $F_{s0} = \mathbf{D}^\ell F_{\ell 0}$.

³ Given a scalar-valued symbol $m(\xi, \eta)$, our definition of the associated bilinear multiplier is

$$\iint e^{ix \cdot (\xi + \eta)} m(\xi, \eta) [\hat{A}_x(\xi), \hat{B}_x(\eta)] \frac{d\xi}{(2\pi)^4} \frac{d\eta}{(2\pi)^4}.$$

For nonzero heat-times s , (2.15) now becomes

$$\mathbf{D}^\alpha F_{\alpha\beta}(s) = w_\alpha, \quad (2.25)$$

where in general w_α , called the *Yang–Mills tension field*, is nontrivial as the two flows (wave and heat) do not commute. Thus additional steps are needed:

(iv) We compute parabolic evolutions for w_α , showing that at time t they depend only on the data $A_x(t)$, $\partial_t A_x(t)$ and of course on s ,

$$w_\alpha = \mathbf{w}_\alpha(A_x(t), \partial_t A_x, s).$$

Moreover, we separate \mathbf{w}_α into an explicit quadratic part and a higher order term,

$$\mathbf{w}_\alpha(s) = \mathbf{w}_\alpha^2(s) + \mathbf{w}_\alpha^3(s)$$

where the latter is purely perturbative.

(v) Finally, we recalculate A_0 and $\mathbf{D}^0 A_0$ to include the dependence on $w(s)$, and write the analogue of the equation (2.19) for $A_x(s)$,

$$\begin{aligned} \square_{A(s)} A_k(s) &= \mathbf{P}[A^j(s), \partial_k A_j(s)] + 2\Delta^{-1} \partial_k \mathbf{Q}(\partial^\alpha A^j(s), \partial_\alpha A_j(s)) + R(A(s), \partial_t A(s)) \\ &\quad + \mathbf{P}\mathbf{w}_k^2(s) + R_s(A, \partial_t A) \end{aligned} \quad (2.26)$$

The extra terms on the right are matched by a like contribution to the quadratic part of A_0 , i.e. (2.20) is replaced by

$$A_0(s) = \mathbf{A}_0^2(A(s), B(s)) + \mathbf{A}_0^3(A(s), B(s)) + \Delta^{-1} \mathbf{w}_0^2(A, B) + \mathbf{A}_{0;s}^3(A, B) \quad (2.27)$$

The s dependent terms in the above equations depend on the original connection A and not just on $A(s)$. However, they have the redeeming feature that they are concentrated at a single dyadic frequency $s^{-\frac{1}{2}}$.

The analysis of the equation (2.26) is now very similar to that of (2.19), with the minor proviso that the quadratic terms in \mathbf{w} in the two equations above have a very mild nonperturbative role, and exhibit a null form type cancellation.

2.6. Remarks on the dynamic Yang–Mills heat flow. In [23], the dynamic Yang–Mills heat flow (2.24) plays a major role in our proofs in several different ways:

- (i) *As a gauge covariant smoothing flow for spacetime connections.* This is the most direct interpretation; (2.24) was used in this capacity to fix the evolution of $w_\mu(s)$ in the preceding subsection.
- (ii) *As a means to perform the “infinitesimal de Turck trick” for the linearized Yang–Mills heat flow in the local caloric gauge.* As alluded to earlier, our understanding of (2.7) is based on its linearization (2.12), which in turn is analyzed through a version of de Turck trick. It implemented as follows, using (2.24) as a useful auxiliary tool:
 - Given a one-parameter family of Yang–Mills heat flows $A_j(t, x, s)$ with data $a_j(t, x)$ ($t \in I, x \in \mathbb{R}^4, s \in J$), we add a t -component $A_0(t, x, s)$ and view it as a connection 1-form on $I \times \mathbb{R}^4 \times J$. In the s -direction, we then *impose* the dynamic Yang–Mills heat flow (2.24).

- Then the key idea is to work with

$$F_{0j} = \partial_t A_j - \mathbf{D}_j A_0. \quad (2.28)$$

As opposed to $\partial_t A_j$, which solves (2.12), F_{0j} has the advantage of obeying a *non-degenerate* covariant parabolic equation:

$$\mathbf{D}_s F_{0j} - \Delta_A F_{0j} - 2ad(F_j^\ell) F_{0\ell} = 0.$$

Solving this equation would determine F_{0j} from any data $F_{0j}(s=0) = e_j$. We choose $e_j = \partial_t a_j$, which amounts to prescribing $a_0 = 0$. Then A_0 may be determined by integrating $\partial_s A_0 = F_{s0} = \mathbf{D}^\ell F_{\ell 0}$, and then we come back to the solution $\partial_t A$ of (2.12).

- (iii) *As a means to obtain useful representation of projection to the caloric manifold.* This is a variant of (2). Previously, we chose to initialize $a_0 = 0$. When $a(t=0)$ is a caloric connection, another natural choice is to set $A_0(s=\infty) = 0$, which amounts to requiring that the nearby $a(t)$'s are also caloric. Integrating $\partial_s A_0 = \mathbf{D}^\ell F_{\ell 0}$ from $s = \infty$ to 0, we obtain

$$a_0 = - \int_0^\infty \mathbf{D}^\ell F_{\ell 0}(s) ds. \quad (2.29)$$

By (2.28), we have

$$e_j = \partial_t a_j - \mathbf{D}_j a_0.$$

Since $a(t)$'s are caloric, $\partial_t a_j$ clearly belongs to $T_a \mathcal{C}$, whereas $\mathbf{D} a_0$ is a pure covariant gradient. This procedure proves Theorem 2.13, while yielding a useful representation formula (2.29).

3. ENERGY DISPERSED CALORIC YANG–MILLS WAVES

Our second article [24] is concerned with the hyperbolic Yang–Mills equation in the caloric gauge, namely the equation (2.19) with the auxiliary variables A_0 and $D_0 A_0$ as in (2.20) and the constraints (2.23).

3.1. Main results in the caloric gauge. The first result is a local well-posedness result which uses the notion of ϵ -energy concentration scale, defined as

$$r_c^\epsilon[a, e] = \sup\{r : \sup_x \int_{B_r(x)} |f|^2 + |e|^2 dx \leq \epsilon^2\}.$$

Then we have

Theorem 3.1 ([24]). *There exists a positive non-increasing function $\epsilon_*(\mathcal{E}, \mathcal{Q})$ so that for any initial data set (a, e) with energy \mathcal{E} and initial caloric size \mathcal{Q} , that the Yang–Mills equation in the caloric gauge is locally well-posed in $\dot{H}^1 \times L^2$ on the time interval $[-r_c^{\epsilon_*}, r_c^{\epsilon_*}]$.*

We omit here the precise meaning of well-posedness, and instead refer the reader to Theorem 5.3 in the last section. Precisely, the conclusions of Theorem 5.3 hold restricted to the interval $[-r_c^{\epsilon_*}, r_c^{\epsilon_*}]$.

The second main result in [24] uses the notion of energy dispersion, first introduced in [32] in the Wave Maps context. For a connection A on a time interval I , we define its *energy dispersion* as

$$\|F\|_{ED[I]} = \sup_k 2^{-2k} \|P_k F\|_{L^\infty L^\infty[I]}.$$

Then we have:

Theorem 3.2. *There exists a positive non-increasing function $\epsilon(\mathcal{E})$ and a nondecreasing function $M(\mathcal{E})$ such that if A is a caloric Yang–Mills wave on I with energy \mathcal{E} and initial caloric size $\mathcal{Q} \lesssim_{\mathcal{E}} 1$ so that $\|F\|_{ED} \leq \epsilon(\mathcal{E})$, then⁴ $\|A\|_{S^1[I]} \leq M(\mathcal{E})$ and A can be continued (as a well-posed solution in the sense of Theorem 3.1) past finite endpoints of I .*

We also note that the initial assumption on \mathcal{Q} only serves to prevent it from being very large. With this assumption, we actually show that $\mathcal{Q}(A) \ll 1$ in the entire interval I . By Theorem 2.4, this assumption can be entirely omitted for subthreshold energies.

These theorems, or rather their contrapositives, can be considered as continuation criteria for the hyperbolic Yang–Mills equation in the caloric gauge. By providing an accurate description of how singularities may occur, they furnish a starting point for the bubble extraction argument in [26], as it will be explained in Section 5.

One downside of using either the caloric gauge (or the Coulomb gauge) is that causality is lost. To remedy this, we prove that the well-posedness property can be transferred from the caloric gauge to the temporal gauge $A_0 = 0$. As a result, we obtain:

Theorem 3.3. *The hyperbolic Yang–Mills equation in \mathbb{R}^{4+1} is globally well-posed in the temporal gauge for all initial data with small energy.*

Unlike the caloric gauge results, however, a downside of Theorem 3.3 is that it does not provide the S^1 regularity of solutions, or any other dispersive bounds.

In the remainder of this section, we will give an overview of ideas in the proofs of Theorems 3.1, 3.2 and 3.3.

3.2. Function spaces. To state the results more precisely, and also to discuss their proof, it is necessary to outline the function spaces framework used in [24], whose main components are the same as in [13, 14]. The core solution space, which we denote by $S^1[I]$, is a Banach space of functions on $I \times \mathbb{R}^4$ with the property that elements of $S^1[I]$ inherit estimates satisfied by free waves in the energy class (i.e., $\square u = 0$ with $(u, \partial_t u)(0) \in \dot{H}^1 \times L^2$), such as energy estimates, Strichartz estimates, (null form) bilinear estimates etc. The corresponding nonlinearity space, denoted by $N[I]$, is defined, on the one hand, small enough to satisfy the inhomogeneous estimate

$$\|u\|_{S^1[I]} \lesssim \|(u, \partial_t u)(0)\|_{\dot{H}^1 \times L^2} + \|\square u\|_{N[I]}, \quad (3.1)$$

and on the other hand, large enough to contain (at least, most of) the nonlinearities of the wave equation (2.19).

Construction of these spaces builds up on many prior works. The space $N[I]$ is simply the sum of the dual energy space (i.e., $L^1 L^2[I]$) and a dual $X^{s,b}$ space. Building blocks of the space $S^1[I]$ include the energy space (i.e., $\|\nabla u\|_{L^\infty L^2[I]}$), the Strichartz spaces (i.e., $\| |D|^{-\alpha} \nabla u \|_{L^p L^q[I]}$ with admissible α, p, q), an $X^{s,b}$ space [5, 1], the refined Strichartz spaces with radial frequency localization [8], and the null frame space [42, 35]. Moreover, we also add a new component S^{sq} (to be described in Section 3.7), which is used in the proof of Theorem 3.3. For the precise definition, we refer to [24, Section 4].

The $S^1[I]$ -norm serves the role of a controlling norm for the caloric Yang–Mills waves. More precisely, we show in [24] that finiteness of this norm implies finer properties of the

⁴The control norm S^1 will be described shortly.

solution itself and those nearby, such as frequency envelope control, persistence of regularity and scattering for A_x , as well as weak Lipschitz dependence and local-in-time continuous dependence for the nearby solutions. For details, see the structure theorems in [24, Section 4].

3.3. Truncated energy dispersion and the central result. It turns out that Theorems 3.1 and 3.2 can be proved essentially at the same time. The idea is to use smallness of the *truncated energy dispersion* at frequencies higher than 2^m ,

$$\|F\|_{ED_{>m}[I]} = \sup_{k>m} 2^{-2k} \|P_k F\|_{L^\infty L^\infty[I]}, \quad (3.2)$$

matched with shortness of the time interval on the scale 2^{-m} . The central result of [24] reads as follows.

Theorem 3.4. *There exist a non-decreasing positive function $M(\mathcal{E}, \mathcal{Q})$ and non-increasing positive functions $\epsilon(\mathcal{E}, \mathcal{Q})$ and $T(\mathcal{E}, \mathcal{Q})$, so that the following holds: For all regular subthreshold caloric Yang–Mills waves A in a time interval I with energy \mathcal{E} and initial caloric size \mathcal{Q} , if we have*

$$\|F\|_{ED_{\geq m}[I]} \leq \epsilon(\mathcal{E}, \mathcal{Q}), \quad |I| \leq 2^{-m} T(\mathcal{E}, \mathcal{Q}), \quad (3.3)$$

then we must also have

$$\|A\|_{S^1[I]} \leq M(\mathcal{E}, \mathcal{Q}). \quad (3.4)$$

On the one hand, this theorem implies an $S^1[I]$ control norm bound on a time interval of size $\leq 2^{-m}$ for data with sufficiently small energy at frequencies $> 2^m$ (i.e., $\|P_{>m}(A_x, \partial_t A_x)(0)\|_{\dot{H}^1 \times L^2}$ is small), which is the case for data with energy concentration scale $\gtrsim 2^{-m}$. On the other hand, it also implies an $S^1[I]$ -bound, independent of I , if the solution has small untruncated energy dispersion $\|F\|_{ED[I]}$. As discussed above, these $S^1[I]$ -norm bounds prove Theorems 3.1 and 3.2, respectively.

3.4. Review of the small energy case: Perturbative nonlinearities and parametrix construction. We begin with a brief discussion of the small energy case, where the goal is to prove $\|A_x\|_{S^1[\mathbb{R}]}^2 \leq C\mathcal{E}$ for sufficiently small \mathcal{E} . This was carried out in [14], which can be viewed as one of the predecessors to this work, in the closely related context of the Coulomb gauge⁵.

The first step was to try to view the wave equation for A_x as a perturbation of the constant coefficient wave equation $\square A_x = 0$. While this is not possible, we can view most of the nonlinearity as perturbative, and estimate them in the space N . In this process, the primary (bilinear) null structure of the Yang–Mills equation, uncovered in [7], plays an essential role. This leaves us with a single nonperturbative term, which arises in a paradifferential fashion,

$$(\square + \text{Diff}_{\mathbf{P}A}^0)A_x := \left(\square + 2 \sum_k \text{ad}(P_{<k} \mathbf{P}^\alpha A) \partial_\alpha P_k \right) A_x = G \quad (3.5)$$

where $\mathbf{P}_x A$ is the Leray projection of A_x , $\mathbf{P}_0 A = A_0$ and G represents a nonlinear but perturbative contribution (which is small thanks to smallness of energy).

Then the key step in [14] was to construct a parametrix for the paradifferential operator $\square + \text{Diff}_{\mathbf{P}A}^0$, and prove that this parametrix satisfies a good $N \rightarrow S^1$ bound akin to (3.1).

⁵While the analysis in [14] is carried out in the Coulomb gauge $\partial^\ell A_\ell = 0$, it is not very different in the caloric gauge, as this also satisfies some form of generalized Coulomb condition $\partial^\ell A_\ell = \mathbf{D}A(A)$.

The rough idea is to try to find a gauge transform O which renormalizes $\square + \text{Diff}_{\mathbf{P}A}^0$ to \square modulo a better behaved error, i.e., schematically

$$(\square + \text{Diff}_{\mathbf{P}A}^0)Ad(O) - Ad(O)\square = (\text{error}), \quad (3.6)$$

and produce a parametrix by conjugating the constant coefficient solution operator by $Ad(O)^{-1}$.

This idea was indeed viable in the case of wave maps [35, 32], but not for Yang–Mills or Maxwell–Klein–Gordon (which may be regarded as a simpler model for Yang–Mills). The difference stems from the structure of the curvature $F[\mathbf{P}A]$, which is a geometric obstruction for gauge transformation of A to 0. Whereas the curvature depends at least quadratically on the solution in the case of wave maps, it is linear (to the leading order) in the solution A for Yang–Mills or Maxwell–Klein–Gordon.

The way out of this difficulty was to consider instead an $Ad(\mathbf{G})$ -valued *pseudodifferential* renormalization operator $Op(Ad(O))$. Heuristically, this generalization allows for separate renormalization of each plane wave solution, which is possible since it only oscillates in a single direction⁶. Using smallness of energy, it was shown that the parametrix obeys the desired $N \rightarrow S^1$, and also that the error in (3.6) is perturbative. We remark that in the error estimate, not only the primary but also the secondary (trilinear) null structure, analogous to that in Maxwell–Klein–Gordon discovered in [16], is crucial.

3.5. Parametrix construction in the large energy case. The difference in the large energy case is that we can no longer use smallness of energy to control neither the perturbative part, nor the parametrix for the paradifferential problem. Thus, in order to be able to close our estimates, we need to have new proxies for smallness.

We start with the paradifferential problem. In departure from the small energy case, but similar to [32, 22], we introduce the *large frequency gap* $\kappa \gg 1$ and consider the paradifferential operator

$$\square + \text{Diff}_{\mathbf{P}A}^\kappa = \square + 2 \sum_k ad(P_{<k-\kappa} \mathbf{P}^\alpha A) \partial_\alpha P_k,$$

where A_x be a caloric Yang–Mills with finite $S^1[I]$ -norm. The goal is to establish an $N \rightarrow S^1$ bound of the form

$$\|u\|_{S^1[I]} \lesssim_{\|A_x\|_{S^1[I]}} \|(u, \partial_t u)(0)\|_{\dot{H}^1 \times L^2} + \|(\square + \text{Diff}_{\mathbf{P}A}^\kappa)u\|_{N[I]}. \quad (3.7)$$

The proof proceeds by a parametrix construction, in a similar manner as [14]. However, the necessary smallness for proving the $N \rightarrow S^1$ bound for the parametrix now comes from taking the frequency gap κ sufficiently large compared to $\|A_x\|_{S^1[I]}$. Moreover, to control the error, we rely on divisibility⁷ of an appropriate weaker norm $\|A_x\|_{DS^1[I]}$ than $\|A_x\|_{S^1[I]}$.

Treating perturbative nonlinearity: Small energy dispersion and short time interval. For the perturbative nonlinearity, smallness may be obtained via truncated energy dispersion and the length of I . Roughly speaking, any unbalanced or close-angle frequency interaction is small (exponentially in the frequency ratio) for such nonlinearities, while balanced and far-angle

⁶This procedure eludes the geometric obstruction mentioned above, since curvature, being a 2-form, always vanishes when restricted to a one-dimensional subspace.

⁷That is, I can be split into a controlled number of subintervals, on each of which the restricted norm is arbitrarily small.

interactions are controlled by $\|F\|_{ED>m[I]}$ at frequencies $\gtrsim 2^m$, and by $2^m|I|$ at frequencies $\lesssim 2^m$. In sum, we have

$$\|F\|_{ED>m[I]} \leq \varepsilon, \quad 2^m|I| \leq \varepsilon \implies \|(\square + \text{Diff}_{\mathbf{P}A}^\kappa)A_x\|_{N[I]} \lesssim_{\|A_x\|_{S^1[I]}} 2^{C\kappa} \varepsilon^\delta.$$

Unfortunately, this bound is insufficient for proving Theorem 3.4. The reason is that the $N \rightarrow S^1$ bound (3.7) for the paradifferential operator already depends on the $S^1[I]$ -norm of A_x , which is what we wish to bound!

3.6. Induction on energy. In order to break the circular argument, we perform an induction on energy, following the scheme developed in [32]. Roughly speaking, the main idea is to view A as a perturbation of another solution \tilde{A} , which has a lower (linear) energy and hence obeys an S^1 -norm bound by an induction hypothesis. To make this idea work, we need to carefully construct \tilde{A} so that we may control the difference $A - \tilde{A}$.

A preliminary step here is to show that \mathcal{Q} is essentially conserved for solutions with small energy dispersion. Once this is done, Q becomes a fixed parameter and is omitted from the subsequent discussion.

The induction argument is set up as follows, in terms of the linear energy E rather than the nonlinear one \mathcal{E} . The initial step is provided by the small energy case, which proves (3.4) up to sufficiently small $E > 0$, with $M(E) = C\sqrt{E}$ and any choices of $\epsilon(E)$, $T(E)$. As the induction hypothesis, we assume that there exist functions $\epsilon(\cdot)$, $T(\cdot)$ and $M(\cdot)$ such that (3.4) holds up to some E . Then the goal is to extend these functions so that (3.4) holds up to $E + c_0$ for some $c_0 = c_0(E) > 0$. An essential point for continuing this induction argument (in order to cover all subthreshold solutions) is to ensure that the increment $c_0(E)$ is *independent* of the functions $\epsilon(\cdot)$, $T(\cdot)$ and $M(\cdot)$ given by the induction hypothesis⁸.

We define \tilde{A} by first flowing the data $\tilde{A}_x(0)$ and $\partial_t \tilde{A}_x(0)$ by the Yang–Mills heat flow and the linearized Yang–Mills heat flow, respectively, for some heat-time s_* , then solving the Yang–Mills equation in caloric gauge in time. Taking ϵ , T and c_0 sufficiently small, and choosing s_* appropriately, we aim for the following two goals:

- i) \tilde{A} exists on I and $\|\tilde{A}\|_{S^1[I]} \leq M(E)$;
- ii) $\|A - \tilde{A}\|_{S^1[I]} \lesssim_{M(E)} 1$.

The cutoff heat-time s_* can be chosen so that either a) $s_* \ll 2^{-m}$ and $\|\nabla \tilde{A}(0)\|_{L^2} = E$, or b) $s_* \simeq 2^{-m}$ and $\|\nabla \tilde{A}(0)\|_{L^2} \geq E$. In both cases, provided that ϵ, T are sufficiently small, it can be shown that \tilde{A}_x is close to the Yang–Mills heat flow $A_x(s_*)$ of A_x . In Case a), taking ϵ smaller if necessary, we may ensure that $\|\tilde{F}\|_{ED \geq m} \leq \epsilon(E)$ and Goal i) follows from the induction hypothesis. In Case b), $\tilde{A}(0)$ is sufficiently smooth so that the desired conclusion can be proved simply by higher order local well-posedness.

To accomplish ii), we need several ideas. First, we observe that the linear energies $\|\nabla A_x(t)\|_{L^2}$, $\|\nabla \tilde{A}_x(t)\|_{L^2}$ of the solutions A, \tilde{A} are conserved in t , up to an error that can be made arbitrarily small by taking ϵ, T small enough. Moreover, since \tilde{A} is close to $A(s_*)$, which in turn is (at least heuristically) a low frequency truncation of A , the frequency supports of $A - \tilde{A}$ and \tilde{A} are essentially separated. Therefore, approximate conservation of linear

⁸Meanwhile, $\epsilon = \epsilon(E + c_0)$, $T = T(E + c_0)$ and $M = M(E + c_0)$ may (and indeed do) depend on $\epsilon(E)$, $T(E)$ and $M(E)$. We are allowed to choose these parameters in the order $c_0 \rightarrow M \rightarrow T, \epsilon$.

energies for A and \tilde{A} implies

$$\sup_{t \in I} \|\nabla(A_x - \tilde{A}_x)(t)\|_{L^2} \lesssim_E \|\nabla A_x(0)\|_{L^2} - \|\nabla \tilde{A}_x(0)\|_{L^2} \leq c_0. \quad (3.8)$$

To upgrade this to an $S^1[I]$ -norm bound, we establish *weak divisibility* of the S^1 -norm of \tilde{A} , i.e., that we can split $I = \cup_{k=1}^K I_k$ so that

$$\|\tilde{A}_x\|_{S^1[I_k]} \lesssim_E 1, \quad K \lesssim_{M(E)} 1. \quad (3.9)$$

Now viewing $A = \tilde{A} + (A - \tilde{A})$ as a perturbation of \tilde{A} on each I_k , where the data for $A - \tilde{A}$ are reinitialized on each interval using (3.8), we may bound the S^1 -norm of $A - \tilde{A}$ on each I_k provided that c_0 is small enough compared to the implicit constants in (3.8) and (3.9). Importantly, these are independent of $M(E)$! Thus Goal B follows by summing up these bounds in $k = 1, \dots, K$.

3.7. Passing to the temporal gauge. Finally, we describe the ideas behind the proof Theorem 3.3. We wish to estimate the gauge transformation O from the caloric gauge into the temporal gauge, which solves the nonlinear transport equation

$$O^{-1} \partial_t O = A_0.$$

For O to preserve \dot{H}^1 regularity of A_x , we need:

$$\Delta A_0 \in \ell^1 L_x^2 L_t^1. \quad (3.10)$$

The proof of (3.10) relies on two observations.

- (i) We note that the following *square function norm* can be added to the S^1 norm, i.e.,

$$\|\nabla A_x\|_{S^{sq}} \lesssim \|A_x\|_{S^1}.$$

where

$$\|u\|_{S^{sq}} = \| |D|^{-\frac{3}{10}} u \|_{\ell^2 L_x^{\frac{10}{3}} L_t^2}.$$

The relevance of $p = \frac{3}{10}$ is that it is the dual Stein–Tomas exponent for Fourier restriction to $\mathbb{S}^3 \subseteq \mathbb{R}^4$. Indeed, the (adjoint) Stein–Tomas restriction theorem and Plancherel in time leads to

$$\|e^{\pm it|D|} u\|_{S^{sq}} \lesssim \|u\|_{L^2},$$

which implies $\nabla u \in S^{sq}$ for \dot{H}^1 free waves. We extend this estimate to our parametrix, which allows us to add S^{sq} into our S^1 norm.

- (ii) In an order zero bilinear expression of the form $\mathbf{O}(A_x, \partial_t A_x)$, the worst case is when $\partial_t A_x$ has the higher frequency. Indeed, the ordinary product $[A_x, \partial_t A_x]$ fails to belong to $\ell^1 L_x^2 L_t^1$ because of this interaction. However, from (2.21), we see that the symbol of $\Delta \mathbf{A}_0^2$ is

$$\Delta \mathbf{A}_0^2(\xi, \eta) = \frac{2|\xi|^2}{|\xi|^2 + |\eta|^2}.$$

which exhibits a favorable gain in the problematic *low* \times *high* interaction!

4. LARGE DATA, CAUSALITY AND THE TEMPORAL GAUGE

Unlike the first two papers, the third one [25] is concerned with large data solutions which are not necessarily topologically trivial, and thus cannot be directly studied using the global caloric gauge. The goal of [25] is two-fold:

- To describe finite energy initial data sets topologically and analytically.
- To use the temporal gauge in order to provide a good local theory for finite energy solutions.

For simplicity we will work in two settings:

- a) For initial data in \mathbb{R}^4 and solutions in \mathbb{R}^{4+1} , or time sections thereof.
- b) For initial data in a ball B_R and solutions in the corresponding uniqueness cone $\mathcal{D}(B_R) = \{|x| + |t| < R\}$ or time sections thereof.

In terms of the initial data, in addition to the energy, a key role is played by the energy concentration scale⁹

$$r_c^\epsilon = \sup\{r > 0 : \mathcal{E}_{B_r(x) \cap X}(a, e) \leq \epsilon \text{ for all } x \in X\},$$

where $X = B_R$ or \mathbb{R}^4 , as well as the outer concentration radius

$$R_c^\epsilon = \inf\{r > 0 : \mathcal{E}_{B(x,r)}[(a, e)] \leq \epsilon \text{ for some } x \in \mathbb{R}^4\}.$$

4.1. Initial data surgery. Here we discuss a technical tool introduced in [25], which may be of independent interest. At various points in the analysis, we need to perform a physical space localization of the Yang–Mills solution. By finite speed of propagation, this task amounts to smoothly cutoff an initial data set (a, e) which turns out to be nontrivial due to the presence of the constraint equation (1.4). To address this issue, we prove the following result:

Theorem 4.1. *Let $B = B_{R_0}(0)$ be a ball centered at 0, and let a be a \dot{H}^1 connection on $\mathbb{R}^4 \setminus B$. Then there exists a solution operator $h \mapsto e = T_a h$ to the equation*

$$\mathbf{D}^\ell e_\ell = h \quad \text{in } \mathbb{R}^4 \setminus B, \tag{4.1}$$

with the following properties:

- (1) *Boundedness: The operator T_a is bounded from \dot{H}^{-1} to L^2 , with a norm depending only on $\|a\|_{L^4}$.*
- (2) *Higher regularity: If a and h are smooth, then $T_a h$ is also smooth.*
- (3) *Exterior support: For any $R \geq R_0$, if $h = 0$ in $B_R(0)$, then $T_a h = 0$ in $B_R(0)$.*

In the case $a = 0$, (4.1) becomes the usual divergence equation and a desired solution operator T_0 may be constructed explicitly. Exploiting the exterior support property of T_0 , T_a is constructed in an essentially inductive manner, starting from an annulus around B (where a can be treated perturbatively) and proceeding outward.

As a quick corollary of Theorem 4.1, we obtain the following initial data excision result.

Proposition 4.2. *Let (a, e) be a small energy data set in $B_4 \setminus B_1$. Then*

- (1) *We can find a small energy exterior data set (\tilde{a}, \tilde{e}) in $\mathbb{R}^4 \setminus B_1$, which agrees with (a, e) in $B_2 \setminus B_1$. Furthermore, if (a, e) is smooth then (\tilde{a}, \tilde{e}) is also smooth.*

⁹For a singlet a , we define r_c^ϵ and R_c^ϵ by taking $e = 0$.

(2) We can find a small energy exterior data set (\tilde{a}, \tilde{e}) in $\mathbb{R}^4 \setminus B_1$, which is gauge equivalent to (a, e) in $B_4 \setminus B_2$. Furthermore, if (a, e) is smooth then (\tilde{a}, \tilde{e}) is also smooth.

The idea of the proof is to first naively extend (a, e) to $\mathbb{R}^4 \setminus B_1$. This generates an error in the constraint equation, which can be removed by applying Theorem 4.1.

Remark 4.3. Theorem 4.1 can clearly be generalized to other regularities and dimensions. In particular, the operator $T_a : \dot{H}^{-1}(\mathbb{R}^3 \setminus B) \rightarrow L^2(\mathbb{R}^3 \setminus B)$ can be used to prove an excision result for finite energy data on \mathbb{R}^3 . We note that this furnishes an alternative approach to constructing local Coulomb gauges [7] that avoids the need to prescribe boundary values.

4.2. Good global gauges. In view of the gauge independence property, having control of the energy of a connection a says little about the $\dot{H}^1 \cap L^4$ size of a . This issue can sometimes be addressed by choosing a good gauge, such as the local Coulomb gauge in Uhlenbeck's Lemma for small energies, or the caloric gauge for subthreshold energies, see Theorems 2.4, 2.8. However, what if our connection has larger energy?

We begin our discussion with initial data sets in a ball. In addition to the energy \mathcal{E} , we also use a second parameter, namely the energy concentration scale $r_c = r_c^\epsilon$, with a small universal constant ϵ . Then we have:

Proposition 4.4. *Given a connection a in B_R with energy \mathcal{E} and energy concentration scale r_c , there exists a gauge equivalent connection \tilde{a} in B_R which satisfies the bound*

$$\|\tilde{a}\|_{\dot{H}^1 \cap L^4} \lesssim_{E, \frac{r_c}{R}} 1 \quad (4.2)$$

Also for initial data in \mathbb{R}^4 , we also can find a good global gauge:

Theorem 4.5 (Good global gauge). *Let $a \in H_{loc}^1(\mathbb{R}^4)$ be a finite energy connection. Then there exists a gauge equivalent representation \tilde{a} of a such that*

$$\tilde{a} = -\chi O_{(\infty);x} + b$$

where $O_{(\infty)}(x)$ is a smooth 0-homogeneous map taking values in \mathbf{G} and $B \in \dot{H}^1$.

Finally, we remark on the relationship between Theorem 4.5 and topological classes of finite energy connections. Precisely, the topological class of a connection a can be parametrized by the homotopy class $[O]$ of the map O in the above theorem, viewed as a map

$$O : \mathbb{S}^3 \rightarrow \mathbf{G}.$$

4.3. The temporal gauge and causality. While we are not able to carry out the full analysis for the Yang–Mills equation in the temporal gauge, we are nevertheless making good use of it in our papers in an auxiliary role. This is due to the following three properties:

- (i) Local well-posedness for regular data.
- (ii) Causality, i.e. finite speed of propagation
- (iii) Agreement with caloric gauge at the linear level.

In our sequence of papers we are taking advantage of these three properties at different places in the analysis. Property (i), for instance, is used in order to prove a local well-posedness for regular data in the caloric gauge, simply by gauge transforming the temporal solutions. Property (iii), essentially described in Section 3.7, allows us to reverse the process, and show that small energy global well-posedness in the caloric gauge implies small energy global well-posedness in the temporal gauge. Finally, as a consequence of property (ii) the

small energy global well-posedness in the temporal gauge implies large energy local well-posedness in the temporal gauge. Even better, it shows that the local solutions can be continued in the temporal gauge for as long as no energy concentration occurs in a light cone.

4.3.1. Finite energy solutions. A consequence of [14] and of the first two papers in the series [23], [24] is that the small data problem for the $4 + 1$ dimensional hyperbolic Yang–Mills equation is well-posed in several gauges: Coulomb, caloric, and temporal. In [25] we exploit the temporal gauge small data result, combined with causality, to obtain results for the large data problem. The local in time result is as follows:

Theorem 4.6 ([25]). *(1) For each finite energy data (a, e) in \mathbb{R}^4 with concentration scale r_c there exists a unique finite energy solution A to (1.1) in the time interval $[-r_c, r_c]$ in the temporal gauge $A_0 = 0$, depending continuously on the initial data. Furthermore, any other finite energy solution with the same data must be gauge equivalent to A .*
(2) The same result holds for data in a ball B_R and the solution in the corresponding domain of uniqueness $\mathcal{D}(C_R) \cap ([-r_c, r_c] \times \mathbb{R}^4)$.

We remark that this caloric gauge well-posedness result is in some sense a soft result, which is not accompanied by any dispersive type estimates. In expanded form, it asserts that regular data generates regular solutions on the r_c time scale, and that the data to solution map has a continuous extension to all finite energy data in the uniform energy norm. However, its proof is anything but straightforward, as it requires the full strength of the local well-posedness in the caloric gauge.

Now we consider the continuation question. The next result asserts that temporal solutions can be continued until energy concentration (i.e. blow-up) occurs. Thus, temporal solutions are also maximal solutions for the Yang–Mills equation.

Theorem 4.7. *(1) For each finite energy data (a, e) in \mathbb{R}^4 , let (T_{min}, T_{max}) be the maximal time interval on which the temporal gauge solution exists. If T_{max} is finite then we have*

$$\lim_{t \rightarrow T_{max}} r_c(t) = 0$$

and similarly for T_{min} . Furthermore, there exists some $X \in \mathbb{R}^4$ so that energy concentrates in the backward light cone of (T_{max}, X) (respectively the forward light cone of (T_{min}, X)).

(2) The same result holds for data in a ball B_R and the solution in the corresponding domain of uniqueness $\mathcal{D}(B_R)$.

The main advantage of this theorem is that it allows us to work with solutions which do not admit a global caloric representation. The vanishing of $r_c(t)$ is a corollary of Theorem 4.6, while existence of a energy concentration point follows by a standard argument; see, e.g., [20, Lemma 8.1].

The temporal gauge is convenient in order to deal with causality, but not so much in terms of regularity, as it lacks good S^1 bounds. For this reason it is convenient to borrow the caloric gauge regularity:

Theorem 4.8. *Let A be a finite energy Yang–Mills solution in a cone section $C_{[t_1, t_2]}$ with energy concentration scale r_c . Then in a suitable gauge A satisfies the bound*

$$\|A\|_{L^\infty(\dot{H}^1 \cap L^4)} + \|\partial_t A\|_{L^\infty L^2} + \|\partial^j A_j\|_{\ell^1 \dot{H}^{\frac{1}{2}}} + \|A_0\|_{\ell^1 \dot{H}^{\frac{3}{2}}} + \|\square A_x\|_{L^2 \dot{H}^{-\frac{1}{2}}} \lesssim_{E, \frac{r_c}{t_2}} 1 \quad (4.3)$$

in the smaller cone $C_{[t_1, t_2]}^{4r_c}$ where the radius has been decreased by $4r_c$.

The proof of this theorem requires a good gluing technique for local connections with suitable regularity, which were used to prove Proposition 4.4 and Theorem 4.5 as well.

5. TO BUBBLE OR NOT TO BUBBLE

In this section we outline the proof of our two main results in Theorems 1.8 and 1.9, following our fourth and the final article [26]. This is based on a blow-up argument based on Morawetz-type monotonicity formulas, broadly following the outline of prior works on Wave Maps [33] and Maxwell–Klein–Gordon [20]. However, new difficulties arise here both at the conceptual level and at the technical level due to the more nonlinear gauge features inherent in Yang–Mills and to the nontrivial topological structure.

We start with a common part to both proofs, namely a energy-based criterion for soliton bubbling-off, and then we consider the two results separately.

5.1. A bubble-off criterion. Our aim here is to describe the proof of the following result, which provides a bubbling-off criterion that applies equally for both the Threshold and the Dichotomy Theorems.

Theorem 5.1 (Bubbling Theorem). *(1) Let A be a finite energy Yang–Mills wave which blows up in finite time at (T, X) . Assume in addition that for some $\gamma < 1$ we have*

$$\limsup_{t \nearrow T} \mathcal{E}_{C_\gamma \cap S_t}(A) > 0, \quad C_\gamma = \{|x - X| \leq \gamma|t - T|\}. \quad (5.1)$$

Then A bubbles off a soliton at (T, X) , as described after Theorem 1.9.

(2) Let A be a finite energy Yang–Mills wave which is global forward in time. Assume in addition that for some $\gamma < 1$ we have

$$\limsup_{t \nearrow \infty} \mathcal{E}_{C_\gamma \cap S_t}(A) > 0, \quad C_\gamma = \{|x| \leq \gamma t\}. \quad (5.2)$$

Then A bubbles off a soliton at infinity, as described after Theorem 1.9.

Beginning of the proof. We start with some notations and initial simplifications. In the finite time blow up case, by translation and reflection we can assume that $(T, X) = (0, 0)$, and that the blow-up occurs in the forward light cone. We introduce the forward cone C , its lateral boundary ∂C and the foliation $\{S_t\}_{t \in [0, \infty)}$ as

$$C = \{(t, x) : 0 \leq |x| \leq t\}, \quad \partial C = \{(t, x) : 0 \leq |x| = t\}, \quad S_t = C \cap (\{t\} \times \mathbb{R}^4).$$

We introduce the energy flux $\mathcal{F}_{[t_1, t_2]}(A)$, defined as

$$\mathcal{F}_{[t_1, t_2]}(A) = \mathcal{E}_{t_2}(A) - \mathcal{E}_{t_1}(A).$$

Assume, for simplicity, that A is regular. Then in both scenarios, by the above energy flux relation, we can easily obtain a sequence $A^{(n)}$ of Yang–Mills waves, all obtained by rescaling the original A , and having the following properties:

- (1) $A^{(n)}$ is defined on $C_{[\varepsilon_n, 1]}$ where $\varepsilon_n \rightarrow 0$;
- (2) (Bounded energy in the cone) $\mathcal{E}_{S_t}(A^{(n)}) \leq E$ for every $t \in [\varepsilon_n, 1]$;
- (3) (Decaying flux on ∂C) $\mathcal{F}_{[\varepsilon_n, 1]}(A^{(n)}) \leq \varepsilon_n^{\frac{1}{2}} E$;
- (4) (Nontrivial time-like energy at $t = 1$) $\mathcal{E}_{C_\gamma \cap S_1}(A^{(n)}) \geq E_0 > 0$.

A Morawetz identity. Here we describe the key monotonicity formula (or a Morawetz identity), from which we obtain both asymptotic stationarity and compactness for bubble extraction. The idea is to use the *renormalized scaling vector field* $X_0 = \frac{1}{\sqrt{t^2 - |x|^2}}(t\partial_t + x \cdot \partial_x)$ as a multiplier. Introducing

$$^{(X_0)}P_\alpha(A) = T_{\alpha\beta}(A)X_0^\beta,$$

where $T_{\alpha\beta}(A)$ is the Yang–Mills energy-momentum tensor, we have

$$\operatorname{div} ^{(X_0)}P(A) = \frac{2}{\rho_0}|\iota_{X_0}F|^2, \quad (5.3)$$

where $\rho_0 = \sqrt{t^2 - |x|^2}$. Remarkably, the RHS is nonnegative!

To derive a monotonicity formula, we would like to integrate (5.3) on $C_{[t_1, t_2]}$ and apply the divergence theorem. However, this is not possible since the weight ρ^{-1} blows up on ∂C . Instead we introduce a parameter $\varepsilon > 0$ and consider $X_\varepsilon = \frac{1}{\rho_\varepsilon}((t + \varepsilon)\partial_t + x \cdot \partial_x)$, where $\rho_\varepsilon = \sqrt{(t + \varepsilon)^2 - |x|^2}$. Introducing the notation

$$^{(X_\varepsilon)}\mathcal{P}_{S_t}(A) = \int_{S_t} ^{(X_0)}P_0(A) dx,$$

we arrive at

$$^{(X_\varepsilon)}\mathcal{P}_{S_{t_2}}(A) + \int_{C_{[t_1, t_2]}} \frac{1}{\rho_\varepsilon}|\iota_{X_\varepsilon}F|^2 dt dx = ^{(X_\varepsilon)}\mathcal{P}_{S_{t_1}} + \int_{\partial C_{[t_1, t_2]}} ^{(X_\varepsilon)}P_\alpha(A)L^\alpha d\text{Area} \quad (5.4)$$

where $L = \partial_t + \frac{x}{|x|} \cdot \partial_x$. In the ideal case when the integral on ∂C vanishes, (5.4) says that the quantity $^{(X_\varepsilon)}\mathcal{P}_{S_t}$ is monotone in t .

To describe $^{(X_\varepsilon)}\mathcal{P}_{S_t}$ in detail, we need more notation. Let $L = \partial_t + \frac{x}{|x|} \cdot \partial_x$, $\underline{L} = \partial_t - \frac{x}{|x|} \cdot \partial_x$, and let $\{e_a\}_{2,3,4}$ be orthonormal vectors which are orthogonal to L, \underline{L} . In terms of the null decomposition of F defined as

$$\alpha_a = F(L, e_a), \quad \underline{\alpha}_a = F(\underline{L}, e_a), \quad \varrho = \frac{1}{2}F(L, \underline{L}), \quad \sigma_{ab} = F(e_a, e_b),$$

we have

$$\begin{aligned} ^{(X_\varepsilon)}\mathcal{P}_{S_t}(A) = \int_{S_t} & \left(\frac{1}{2} \left(\frac{t+r+\varepsilon}{t-r+\varepsilon} \right)^{1/2} (|\alpha|^2 + |\varrho|^2 + |\sigma|^2) \right. \\ & \left. + \frac{1}{2} \left(\frac{t-r+\varepsilon}{t+r+\varepsilon} \right)^{1/2} (|\underline{\alpha}|^2 + |\varrho|^2 + |\sigma|^2) \right) dx. \end{aligned} \quad (5.5)$$

Finally, we discuss how (5.4) is applied to our setting. For the solution $A^{(n)}$ constructed above, the RHS of (5.4) can be bounded by $\lesssim E$ for $\varepsilon = \varepsilon_n$. We point out that the last term is bounded by the energy flux $\mathcal{F}_{[t_1, t_2]}(A)$. Thus

$$\sup_{t \in (\varepsilon_n, 1]} ^{(X_{\varepsilon_n})}\mathcal{P}_{S_t}(A^{(n)}) + \iint_{C_{(\varepsilon_n, 1]}} \frac{1}{\rho_{\varepsilon_n}}|\iota_{X_{\varepsilon_n}}F^{(n)}|^2 dt dx \lesssim E. \quad (5.6)$$

Consider a time-like cone $C_\gamma = \{(t, x) : |x| \leq \gamma t\}$ for any $0 < \gamma < 1$. Observe that $\rho_\varepsilon \simeq t$ and X_ε is uniformly time-like in $C_\gamma \cap \{t \geq 2\varepsilon\}$ (both statements are uniform as $\varepsilon \rightarrow 0$ but degenerate as $\gamma \rightarrow 1$). Thus boundedness of the spacetime integral term in (5.6) implies logarithmic integrated decay of a uniformly time-like interior derivative of $F^{(n)}$ in C_γ ; this decay is the source of asymptotic stationarity and compactness.

Propagating energy in time-like region. The monotonicity formula (5.4) suggests that the weighted energy $^{(X_0)}\mathcal{P}_{S_t}(A^{(n)})$ essentially increases toward the tip. Using a suitably localized version of the formula, we show that nontrivial energy persists in a time-like cone toward the tip:

$$\mathcal{E}_{C_\gamma \cap S_t}(A^{(n)}) \geq E_1 \quad \text{for } t \in [\varepsilon_n^{\frac{1}{2}}, \varepsilon_n^{\frac{1}{4}}], \quad (5.7)$$

where we make $1 - \gamma$ and E_1 smaller if necessary.

Final rescaling. After a pigeonhole argument and suitable rescalings, we obtain a sequence of caloric Yang–Mills waves on $[1, T_n] \times \mathbb{R}^4$ (where $T_n \rightarrow \infty$), which we still denote by A^n , with the following properties (final rescaled sequence):

- (1) (Bounded energy in the cone) $\mathcal{E}_{S_t}(A^{(n)}) \leq E \quad (t \in [1, T_n]);$
- (2) (Small energy outside the cone) $\mathcal{E}_{(\{t\} \times \mathbb{R}^4) \setminus S_t}(A^{(n)}) \ll E \quad (t \in [1, T_n]);$
- (3) (Nontrivial energy in a time-like region) $\mathcal{E}_{C_\gamma \cap S_t}(A^{(n)}) \geq E_1 \quad (t \in [1, T_n]);$
- (4) (Asymptotic self-similarity) For every compact subset \tilde{C} of $C_{[1, \infty)}^1 = \{(t, x) \in C : |x - |t|| \geq 1\}$,

$$\iint_{\tilde{C}} |\iota_{X_0} F^{(n)}| dt dx \rightarrow 0 \quad \text{as } n \rightarrow \infty. \quad (5.8)$$

Locating concentration scales. To extract a bubble, we now locate (locally) smallest concentration scales in $A^{(n)}$, which retains the decay (5.8). A combinatorial argument from [20] (based on [33]) establishes two possible scenarios (along a subsequence of $A^{(n)}$):

- i) (Time-like concentration) There exists $r > 0$, a sequence of points $(t_n, x_n) \rightarrow (t_0, x_0) \in \text{Int}(C_{[1, \infty)})$, and a sequence of scales $r_n \rightarrow 0$ such that

$$\sup_{x \in B_r(x_n)} \mathcal{E}_{B_{r_n}(x)}(A^{(n)})$$

is uniformly small but nontrivial, yet

$$\frac{1}{2r_n} \int_{t_n - r_n}^{t_n + r_n} \int_{B_r(x_n)} |\iota_V F^{(n)}| dt dx \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

where $V = X_0(t_0, x_0)$.

- ii) (Self-similar concentration) For every set¹⁰ of the form

$$\tilde{C} = \{(t, x) : 0 \leq |x| < t - \frac{1}{2}, 2^j \leq t < 2^{j+1} \text{ for some } j \in \mathbb{Z}\}$$

there exists $r = r(\tilde{C})$ such that

$$\sup_{x \in \tilde{C}} \mathcal{E}_{B_r(x)}(A^{(n)})$$

is uniformly small.

¹⁰In fact, any compact subset \tilde{C} in the interior of $C_{[1, \infty)}^1$ would work

Local compactness result. In both scenarios, we would like to extract a limit modulo scalings, translations and gauge transformations. To ensure that the limit is nontrivial and solves the hyperbolic Yang–Mills equation, we need a means to ensure compactness.

Theorem 5.2. *Let $A^{(n)}$ be a sequence of finite energy Yang–Mills connections in $[-2, 2] \times \mathbb{R}^4$ which is locally uniformly bounded in the sense of (4.3). Let $Q = [-1, 1] \times B_R(0)$ and $2Q = [-2, 2] \times B_{2R}(0)$. Assume that*

$$\lim_{n \rightarrow \infty} \|\iota_X F\|_{L^2(2Q)} = 0,$$

where X is a smooth time-like vector field. Then on a subsequence, we have

$$A^{(n)} \rightarrow A \quad \text{in } H^1(Q),$$

where A is a solution to the Yang–Mills equation satisfying $\iota_X F = 0$.

The idea of the proof is as follows. The S^1 bound implies uniform boundedness of $\|\square A^{(n)}\|_{L^2 \dot{H}^{-\frac{1}{2}}}$. This in turn implies extra regularity away from the characteristic cone $\{|\tau| = |\xi|\}$ in frequency space, since \square is elliptic there. Near the characteristic cone, we use the following equation for $A^{(n)}$:

$$\begin{aligned} X^\alpha \partial_\alpha A_j^{(n)} - X^\ell \partial_j A_\ell^{(n)} &= -(\iota_X F^{(n)})_j + (\text{smoother error}), \\ X^\ell \partial_0 A_\ell^{(n)} &= -(\iota_X F^{(n)})_0 + (\text{smoother error}). \end{aligned}$$

Although the system on the LHS is not elliptic, it is microlocally elliptic (of order 1) near the characteristic cone $\{|\tau| = |\xi|\}$ in frequency space. Inverting this system, and using the hypothesis $\iota_X F^{(n)} \rightarrow 0$ in $L^2(2Q)$, we arrive at the decomposition

$$A^{(n)} = A^{(n), \text{small}} + A^{(n), \text{smooth}}, \quad \|A^{(n), \text{small}}\|_{H^1(Q)} \rightarrow 0, \quad \|A^{(n), \text{smooth}}\|_{H^{1+\alpha}(Q)} \lesssim 1,$$

for some $\alpha > 0$ (in fact, $\alpha = \frac{1}{2}$). Applying Rellich–Kondrachov to $A^{(n), \text{smooth}}$, the theorem follows.

Extraction of limiting profiles. In order to apply Theorem 5.2 in Scenario i), we rescale and translate so that $B_{r_n}(x_n) \rightarrow B_1(0)$ and apply Theorem 4.8 to insure the bound (4.3), uniformly on bounded sets. As a result, we extract a nontrivial finite energy stationary solution (i.e., a soliton).

In Scenario ii), we apply a similar procedure to $B_r(0)$, where we rely on Property (4) of the final rescaled sequence for the decay hypothesis in Theorem 5.2. In this case, we extract a finite energy self-similar solution on $C_{[1, \infty)}^1$, which is nontrivial thanks to Property (3).

Exclusion of the self-similar case. To conclude the bubble extraction argument, it remains to rule out Scenario ii), i.e., to prove that every finite energy self-similar solution is trivial.

By self-similarity, the solution restricted to the hyperbolic space $\mathbb{H}^4 = \{(t, x) : t > 0, t^2 - |x|^2 = 1\}$ is a harmonic Yang–Mills connection. Recall that the harmonic Yang–Mills equation in dimension 4 is conformally invariant. Thus, by a stereographic projection, we obtain a harmonic Yang–Mills connection on \mathbb{D}^4 , which we still denote by A . Finite energy condition restricted to the hyperbolic space \mathbb{H}^4 essentially implies that, after a suitable gauge transformation, A is smooth up to the boundary and $A|_{\partial \mathbb{D}^4}$ vanishes. By an elliptic unique continuation argument (applied to F), it follows that the solution is trivial.

5.2. The Threshold Theorem. We first restate our Threshold Theorem in the caloric gauge. We will consider the global solvability question for the system (1.1) with initial data at time $t = 0$

$$(A_j(0), \partial_0 A_j(0)) = (A_{0j}, B_{0j}) \in T^{L^2} \mathcal{C} \subset \mathcal{H} := \mathbf{H}(\mathbb{R}^4) \times L^2(\mathbb{R}^4). \quad (5.9)$$

Here the caloric gauge imposes a constraint on both A_{0j} and on B_{0j} . As discussed before, the temporal components of the connection, namely A_0 and $\partial_0 A_0$, are determined in an elliptic fashion in terms of A_x and $\partial_0 A_x$.

We will also consider higher regularity and (weak) Lipschitz dependence properties of the solutions, using the spaces

$$\mathcal{H}^\sigma = \dot{\mathcal{H}}^\sigma \cap \mathcal{H}, \quad \dot{\mathcal{H}}^\sigma = \dot{H}^\sigma(\mathbb{R}^4) \times \dot{H}^{\sigma-1}(\mathbb{R}^4).$$

Now we can provide a more complete statement for our main result:

Theorem 5.3. *The Yang–Mills system in the caloric gauge (1.1) is globally well-posed in \mathcal{H} for all caloric initial data in \mathcal{H} below the ground state energy, in the following sense:*

(i) *(Regular data) If in addition the data set (A_{0j}, B_{0j}) is more regular, $(A_{0j}, B_{0j}) \in \mathcal{H}^N$, then there exists a unique global regular caloric solution $(A_j, \partial_0 A_j) \in C(\mathbb{R}, \mathcal{H}^N)$, also with $(A_0, \partial_0 A_0) \in C(\mathbb{R}, \mathcal{H}^N)$, which has Lipschitz dependence on the initial data locally in time in the \mathcal{H}^N topology.*

(ii) *(Rough data) The flow map admits an extension*

$$T^{L^2} \mathcal{C} \ni (A_{0j}, B_{0j}) \rightarrow (A_\alpha, \partial_t A_\alpha) \in C(\mathbb{R}, \mathcal{H})$$

and which is continuous in the $\mathcal{H} \cap \dot{\mathcal{H}}^s$ topology for $s < 1$ and close to 1.

(iii) *(Weak Lipschitz dependence) The flow map is globally Lipschitz in the $\dot{\mathcal{H}}^s$ topology for $s < 1$, close to 1.*

We remark that in effect the proof of the theorem provides a stronger statement, where the regularity of the solutions is described in terms of function spaces S^1 , S^N which incorporate, in particular, Strichartz norms, $X^{s,b}$ norms and null frame spaces.

Implicit in Theorem 5.3 is also a scattering result; however, this is not so easy to state as it is a modified rather than linear scattering. In a weaker sense, one can think of scattering as simply the fact that the S^1 norm is finite.

In what follows we outline the proof, using Theorems 3.1, 3.2 and 5.1 as our starting point.

5.2.1. No bubbling. The first step here is to show that no bubbling can occur. Here, we closely follow the argument in [15].

Indeed, assume by contradiction that a sequence $A^{(n)}$ of rescales and translates of A converges locally in H^1 to a Lorentz transform of a nontrivial soliton $L_v Q$, which implies L_{loc}^2 convergence of curvature tensors $F^{(n)}$. So after taking a subsequence, for almost every t

$$\mathcal{E}_{\{t\} \times B_R}(A^{(n)}) = \frac{1}{2} \int_{B_R} \langle F^{(n)}, F^{(n)} \rangle(t) \rightarrow \mathcal{E}_{\{t\} \times B_R}(L_v Q) \quad \text{for any } R > 0,$$

which in turn implies

$$\mathcal{E}(Q) \leq \mathcal{E}(A) < 2E_{\text{GS}}.$$

By Theorem 1.7, the only possibility for Q is that $|\chi(Q)| = \mathcal{E}_e(Q)$. Moreover, since Lorentz transforms preserve the topological class, $\chi(L_v(Q)) = \chi(Q)$.

By topological triviality of $A^{(n)}(t)$, we have $\chi(A^{(n)}(t)) = 0$, and thus

$$\int_{\mathbb{R}^4 \setminus B_R(0)} -\langle F^{(n)} \wedge F^{(n)} \rangle(t) = - \int_{B_R(0)} -\langle F^{(n)} \wedge F^{(n)} \rangle(t).$$

By L_{loc}^2 convergence of $F^{(n)}$, the absolute value of the first term on the RHS can be made arbitrarily close to $|\chi(A)| = \mathcal{E}(Q)$ by taking R very large. Using the Bogomoln'yi lower bound $|\langle F \wedge F \rangle| \leq \frac{1}{2} \langle F_{ij}, F^{ij} \rangle$ in $\mathbb{R}^4 \setminus B_R$, it follows that

$$\begin{aligned} \mathcal{E}(A) &\geq \limsup_{n \rightarrow \infty} \left(\frac{1}{2} \int_{B_R} \langle F^{(n)}, F^{(n)} \rangle(t) + \left| \int_{\mathbb{R}^4 \setminus B_R} \langle F^{(n)} \wedge F^{(n)} \rangle(t) \right| \right) \\ &\geq \mathcal{E}_{\{t\} \times B_R}(L_v Q) + \left| \int_{B_R} -\langle F[L_v Q] \wedge F[L_v Q] \rangle \right| \\ &\geq \mathcal{E}(L_v Q) + \mathcal{E}(Q) - o_{R \rightarrow \infty}(1). \end{aligned}$$

Since $\mathcal{E}(L_v Q) \geq \mathcal{E}(Q) \geq E_{\text{GS}}$, we reach a contradiction.

5.2.2. No blow-up. Suppose finite time blow-up occurs for a subthreshold caloric Yang–Mills wave. By translation invariance we can assume that the blow-up happens at $(0, 0)$, backwards in time. By the small data result, we must have energy concentration in the forward light cone C at $t = 0$

$$\lim_{t \searrow 0} \mathcal{E}_{S_t}(A) > 0. \quad (5.10)$$

On the other hand, as bubbling cannot occur, by Theorem 5.1 we must have

$$\lim_{t \searrow 0} \mathcal{E}_{C_\gamma \cap S_t}(A) = 0 \quad \forall \gamma < 1. \quad (5.11)$$

To reach a contradiction, it would suffice to show that the energy dispersion decays near the tip of the cone,

$$\lim_{t \searrow 0} \|F\|_{ED[0,t]} = 0.$$

Then Theorem 3.2 would yield a bound for $\|A\|_{S^1[0,t]}$, which shows that the solution A extends below $t = 0$ and in particular the energy concentration (5.10) cannot occur.

One problem with this strategy is that we have no a-priori knowledge about what happens outside the cone. To rectify this we excise the outer part of the solution, so that we are left with a connection \tilde{A} in a small time interval $[0, t_0]$, so that

- (1) The two connections agree inside, $\tilde{A} = A$ in $C_{[0,t_0]}$.
- (2) \tilde{A} has small energy outside,

$$\mathcal{E}_{\mathbb{R}^4 \setminus C_t}(\tilde{A}) \leq \epsilon \ll 1, \quad t \in [0, t_0] \quad (5.12)$$

Here ϵ can be chosen arbitrarily small, and t_0 depends on ϵ . This is achieved using Proposition 4.2 at a well chosen time t_0 , using the flux decay near the tip of the cone. By finite speed of propagation, note that the new and old solutions agree in C . In particular, the new solution also concentrates energy at $(0, 0)$, and thus cannot be extended past 0.

Taking into account (5.12) and (5.11) (the latter transfers from A to \tilde{A}) for \tilde{A} , we see that the energy of \tilde{A} must concentrate near the cone. Using the Morawetz estimate (5.6), we obtain as well a second energy bound inside the cone, namely

$$\limsup_{t \rightarrow 0} {}^{(X^\gamma)} \mathcal{P}_{S_t}[\tilde{A}] \lesssim_{\mathcal{E}} 1, \quad \gamma < 1. \quad (5.13)$$

This shows that in addition, only certain curvature components may be large near the cone.

Finally, we are in a position to show that \tilde{A} is energy dispersed near the tip, and thus reach the desired contradiction by Theorem 3.2. This is done using the following result:

Proposition 5.4. *Let $(A_x, \partial_t A_x)(t)$ be a caloric Yang–Mills data with energy $\mathcal{E} < 2E_{GS}$. Then for each $\epsilon > 0$ there exists $\gamma < 1$ and $\delta > 0$ so that the bounds*

$$\mathcal{E}_{C_\gamma \cap S_t(A)}(A) + \mathcal{E}_{\mathbb{R}^4 \setminus S_t}(\tilde{A}) \leq \delta, \quad (X^\gamma) \mathcal{P}_{S_t}[\tilde{A}] \lesssim_\epsilon 1$$

imply

$$\|F\|_{ED[t]} \leq \epsilon.$$

Indeed, by the huge weight near ∂C in $(X_{\varepsilon_n}) \mathcal{P}_{S_1}(A)$ and smallness of energy elsewhere, all components of F except for $\underline{\alpha}$ are small in L^2 . To control $\underline{\alpha}$, it suffices to consider $F_{ra} = \alpha_a - \underline{\alpha}_a$ in the frame $(e_t = \partial_t, e_r = \partial_r, e_2, e_3, e_4)$. By the Yang–Mills equation and the Bianchi identity, they obey the following covariant div-curl system on spheres¹¹:

$$\begin{aligned} \mathbf{D}_a F_{rb} - \mathbf{D}_b F_{ra} &= \mathbf{D}_r \sigma_{ab}, \\ \mathbf{D}^a F_{ra} &= \mathbf{D}^a \alpha_a + \mathbf{D}_r \varrho. \end{aligned}$$

The crucial observation is that the RHS only involve components with small energy. In the abelian case (where $\mathbf{D} = \nabla$), this div-curl system can be easily inverted, and it follows that $\| |\nabla_x|^{-1} \nabla F_{ra} \|_{L^2} \ll E$, where $\nabla = (\nabla_{e_2}, \nabla_{e_3}, \nabla_{e_4})$ stands for the angular derivatives. By Bernstein, this is sufficient to rule out the null concentration scenario. A more involved argument is needed in the non-abelian case.

5.2.3. Scattering. The argument here is similar but simpler. Simply by translating the coordinate system we can insure that the condition (5.12) holds for $t \in [t_0, \infty)$. Then the rest of the argument carries through unchanged.

5.3. The Dichotomy Theorem. Here we would like to apply the same argument as before. This time we are assuming, rather than proving that bubbling does not happen. We can still truncate the solution A outside to insure that the bound (5.12) holds in the blow-up case, or translate the coordinates to achieve the same outcome in the non-scattering case. The new difficulty is that we are no longer guaranteed that we can work in the caloric gauge, as the energy may be above the threshold.

However, it turns out that this is only a technical obstruction, as we can now prove a much stronger form of

Proposition 5.5. *Let $(A_x, F_{0x})(t)$ be a finite energy Yang–Mills data with energy \mathcal{E} . Then for each $\epsilon > 0$ there exists $\gamma < 1$ and $\delta > 0$ so that the bounds*

$$\mathcal{E}_{C_\gamma \cap S_t(A)}(A) + \mathcal{E}_{\mathbb{R}^4 \setminus S_t}(\tilde{A}) \leq \delta, \quad (X^\gamma) \mathcal{P}_{S_t}[\tilde{A}] \lesssim_\epsilon 1$$

imply that $(A_x, F_{0x})(t)$ admits a caloric gauge representation so that in addition we have

$$\|F\|_{ED[t]} \leq \epsilon$$

The difficulty here is to obtain the caloric gauge representation, without assuming any a-priori bound on $\|A[t]\|_{\dot{H}^1 \times L^2}$. This is done via multiple continuity arguments, in several steps:

¹¹We remark that in our actual proof, we work with an analogous div-curl system on hyperplanes for technical simplicity.

- (i) Working in an annulus, use a continuity argument show that one can obtain a local gauge which where A is controlled in \dot{H}^1 , with small L^4 norm.
- (ii) Extend previous step to all of \mathbb{R}^4 , by gluing small $\dot{H}^1 \cap L^4$ connections obtained via Uhlenbeck's lemma inside the annulus and outside.
- (iii) Use a second continuity argument to show that a corresponding caloric connection exists. Here the previous step is used to construct a path to 0.

REFERENCES

- [1] J. Bourgain, *Global solutions of nonlinear Schrödinger equations*, American Mathematical Society Colloquium Publications, vol. 46, American Mathematical Society, Providence, RI, 1999, [doi](#).
- [2] D. M. Eardley and V. Moncrief, *The global existence of Yang-Mills-Higgs fields in 4-dimensional Minkowski space. I. Local existence and smoothness properties*, Comm. Math. Phys. **83** (1982), no. 2, 171–191, [link](#).
- [3] ———, *The global existence of Yang-Mills-Higgs fields in 4-dimensional Minkowski space. II. Completion of proof*, Comm. Math. Phys. **83** (1982), no. 2, 193–212, [link](#).
- [4] M. Gursky, C. Kelleher, and J. Streets, *A conformally invariant gap theorem in Yang-Mills theory*, preprint (2017), [arXiv:1708.01157](#).
- [5] S. Klainerman and M. Machedon, *Space-time estimates for null forms and the local existence theorem*, Comm. Pure Appl. Math. **46** (1993), no. 9, 1221–1268, [doi](#).
- [6] ———, *On the Maxwell-Klein-Gordon equation with finite energy*, Duke Math. J. **74** (1994), no. 1, 19–44, [doi](#).
- [7] ———, *Finite energy solutions of the Yang-Mills equations in \mathbb{R}^{3+1}* , Ann. of Math. (2) **142** (1995), no. 1, 39–119, [doi](#).
- [8] S. Klainerman and D. Tataru, *On the optimal local regularity for Yang-Mills equations in \mathbb{R}^{4+1}* , J. Amer. Math. Soc. **12** (1999), no. 1, 93–116, [doi](#).
- [9] J. Krieger, W. Schlag, and D. Tataru, *Renormalization and blow up for the critical Yang-Mills problem*, Adv. Math. **221** (2009), no. 5, 1445–1521, [doi](#).
- [10] J. Krieger and J. Lührmann, *Concentration compactness for the critical Maxwell-Klein-Gordon equation*, Ann. PDE **1** (2015), no. 1, Art. 5, 208.
- [11] J. Krieger and W. Schlag, *Concentration compactness for critical wave maps*, EMS Monographs in Mathematics, European Mathematical Society (EMS), Zürich, 2012, [doi](#).
- [12] J. Krieger and J. Sterbenz, *Global regularity for the Yang-Mills equations on high dimensional Minkowski space*, Mem. Amer. Math. Soc. **223** (2013), no. 1047, vi+99, [doi](#).
- [13] J. Krieger, J. Sterbenz, and D. Tataru, *Global well-posedness for the Maxwell-Klein-Gordon equation in $4+1$ dimensions: small energy*, Duke Math. J. **164** (2015), no. 6, 973–1040, [doi](#).
- [14] J. Krieger and D. Tataru, *Global well-posedness for the Yang-Mills equation in $4+1$ dimensions. Small energy*, Ann. of Math. (2) **185** (2017), no. 3, 831–893.
- [15] A. Lawrie and S.-J. Oh, *A refined threshold theorem for $(1+2)$ -dimensional wave maps into surfaces*, Comm. Math. Phys. **342** (2016), no. 3, 989–999, [doi](#).
- [16] M. Machedon and J. Sterbenz, *Almost optimal local well-posedness for the $(3+1)$ -dimensional Maxwell-Klein-Gordon equations*, J. Amer. Math. Soc. **17** (2004), no. 2, 297–359, [doi](#).
- [17] V. Moncrief, *Global existence of Maxwell-Klein-Gordon fields in $(2+1)$ -dimensional spacetime*, J. Math. Phys. **21** (1980), no. 8, 2291–2296, [doi](#).
- [18] S.-J. Oh, *Gauge choice for the Yang-Mills equations using the Yang-Mills heat flow and local well-posedness in H^1* , J. Hyperbolic Differ. Equ. **11** (2014), no. 1, 1–108, [doi](#).
- [19] ———, *Finite energy global well-posedness of the Yang-Mills equations on \mathbb{R}^{1+3} : an approach using the Yang-Mills heat flow*, Duke Math. J. **164** (2015), no. 9, 1669–1732, [doi](#).
- [20] S.-J. Oh and D. Tataru, *Global well-posedness and scattering of the $(4+1)$ -dimensional Maxwell-Klein-Gordon equation*, Invent. Math. **205** (2016), no. 3, 781–877, [arXiv:1503.01562](#), [doi](#).
- [21] ———, *Local well-posedness of the $(4+1)$ -dimensional Maxwell-Klein-Gordon equation at energy regularity*, Ann. PDE **2** (2016), no. 1, Art. 2, 70, [arXiv:1503.01560](#), [doi](#).

- [22] ———, *Energy dispersed solutions for the $(4+1)$ -dimensional Maxwell-Klein-Gordon equation*, Amer. J. Math. (2017), [arXiv:1503.01561](#).
- [23] ———, *The Yang-Mills heat flow and the caloric gauge*, preprint (2017).
- [24] ———, *The hyperbolic Yang-Mills equation in the caloric gauge. Local well-posedness and control of energy dispersed solutions*, preprint (2017).
- [25] ———, *The hyperbolic Yang-Mills equation for connections in an arbitrary topological class*, preprint (2017).
- [26] ———, *The Threshold Conjecture for the energy critical hyperbolic Yang-Mills equation*, preprint (2017).
- [27] P. Raphaël and I. Rodnianski, *Stable blow up dynamics for the critical co-rotational wave maps and equivariant Yang-Mills problems*, Publ. Math. Inst. Hautes Études Sci. **115** (2012), 1–122, [doi](#).
- [28] I. Rodnianski and T. Tao, *Global regularity for the Maxwell-Klein-Gordon equation with small critical Sobolev norm in high dimensions*, Comm. Math. Phys. **251** (2004), no. 2, 377–426, [doi](#).
- [29] A. Schlatter, *Long-time behaviour of the Yang-Mills flow in four dimensions*, Ann. Global Anal. Geom. **15** (1997), no. 1, 1–25, [doi](#).
- [30] S. Selberg and A. Tesfahun, *Null structure and local well-posedness in the energy class for the Yang-Mills equations in Lorenz gauge*, J. Eur. Math. Soc. (JEMS) **18** (2016), no. 8, 1729–1752, [doi](#).
- [31] L. M. Sibner, R. J. Sibner, and K. Uhlenbeck, *Solutions to Yang-Mills equations that are not self-dual*, Proc. Nat. Acad. Sci. U.S.A. **86** (1989), no. 22, 8610–8613, [doi](#).
- [32] J. Sterbenz and D. Tataru, *Energy dispersed large data wave maps in $2 + 1$ dimensions*, Comm. Math. Phys. **298** (2010), no. 1, 139–230, [doi](#).
- [33] ———, *Regularity of wave-maps in dimension $2 + 1$* , Comm. Math. Phys. **298** (2010), no. 1, 231–264, [doi](#).
- [34] M. Struwe, *The Yang-Mills flow in four dimensions*, Calc. Var. Partial Differential Equations **2** (1994), no. 2, 123–150, [doi](#).
- [35] T. Tao, *Global regularity of wave maps. II. Small energy in two dimensions*, Comm. Math. Phys. **224** (2001), no. 2, 443–544, [doi](#).
- [36] ———, *Geometric renormalization of large energy wave maps*, Journées “Équations aux Dérivées Partielles”, École Polytech., Palaiseau, 2004, pp. Exp. No. XI, 32.
- [37] ———, *Global regularity of wave maps III. Large energy from βR^{1+2} to hyperbolic spaces*, preprint (2008), [arXiv:0805.4666](#).
- [38] ———, *Global regularity of wave maps IV. Absence of stationary or self-similar solutions in the energy class*, preprint (2008), [arXiv:0806.3592](#).
- [39] ———, *Global regularity of wave maps V. Large data local wellposedness and perturbation theory in the energy class*, preprint (2008), [arXiv:0808.0368](#).
- [40] ———, *Global regularity of wave maps VI. Abstract theory of minimal-energy blowup solutions*, preprint (2009), [arXiv:0906.2833](#).
- [41] ———, *Global regularity of wave maps VII. Control of delocalised or dispersed solutions*, preprint (2009), [arXiv:0908.0776](#).
- [42] D. Tataru, *On global existence and scattering for the wave maps equation*, Amer. J. Math. **123** (2001), no. 1, 37–77, [link](#).
- [43] ———, *Rough solutions for the wave maps equation*, Amer. J. Math. **127** (2005), no. 2, 293–377, [link](#).
- [44] K. K. Uhlenbeck, *Connections with L^p bounds on curvature*, Comm. Math. Phys. **83** (1982), no. 1, 31–42, [link](#).
- [45] ———, *Removable singularities in Yang-Mills fields*, Comm. Math. Phys. **83** (1982), no. 1, 11–29, [link](#).

KIAS, SEOUL, KOREA 02455
E-mail address: sjoh@kias.re.kr

DEPARTMENT OF MATHEMATICS, UC BERKELEY, BERKELEY, CA, 94720
E-mail address: tataru@math.berkeley.edu